

BAYESIAN ANALYSIS OF ITEM RESPONSE MODELS FOR BINARY DATA

By

ATALANTA GHOSH

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

1996

To my parents and teachers

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professors Malay Ghosh and Alan Agresti for being my advisors. Without their enormous patience, encouragement and guidance, it would not have been possible to complete the work. I consider myself extremely lucky to get them as my dissertation advisors. I would like to thank my friends for their great help, especially during the final stage of this dissertation. Also I would like to wholeheartedly thank my wife Sofia Paul for her invaluable support and help all throughout the work.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	vi
CHAPTERS	
1 INTRODUCTION	1
1.1 Literature Review	1
1.2 Topic of this Dissertation	6
2 A UNIFIED BAYESIAN ANALYSIS OF ITEM RESPONSE MODELS FOR BINARY DATA	8
2.1 Introduction	8
2.2 Choice Of Priors	9
2.3 Implementation Of Bayes Procedures	23
2.4 Analysis Based on Data For Placement Tests	26
2.5 Example on Matched Pairs Data	31
2.5.1 Effect of Diagonal Elements	39
2.6 Uniform approximation of improper priors by proper priors	44
2.7 Consistency of Marginal Maximum Likelihood Estimator	57
3 HIERARCHICAL BAYESIAN ANALYSIS OF ITEM RESPONSE MODELS FOR BINARY DATA	75
3.1 Introduction	75
3.2 Bayes Procedures with Hierarchical Priors	76
3.3 Implementation Of Bayes Procedures	81
3.4 An Example	82
4 A UNIFORM BAYESIAN ANALYSIS OF TWO-PARAMETER ITEM RE- SPONSE MODELS FOR BINARY DATA	86
4.1 Introduction	86

4.2	Choice Of Priors	87
4.3	Implementation Of Bayes Procedures	94
4.4	An Example	96
5	SUMMARY AND FUTURE RESEARCH	101
	BIBLIOGRAPHY	104
	BIOGRAPHICAL SKETCH	108

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment
of the Requirements for the Degree of
Doctor of Philosophy

BAYESIAN ANALYSIS OF ITEM RESPONSE MODELS FOR BINARY DATA

By

Atalanta Ghosh

December 1996

Chairpersons: Alan Agresti Malay Ghosh
Major Department: Statistics

We present a unified Bayesian approach for the analysis of one- and two-parameter item response models, with special emphasis on logit, probit, and log-log links. Necessary and sufficient conditions are found for the propriety of the posteriors under improper priors for these models. Bayes estimation is implemented using Markov Chain Monte Carlo integration technique, and is illustrated with an example from educational testing. Some relationships between the frequentist and the Bayesian procedure are also discussed.

CHAPTER 1

INTRODUCTION

1.1 Literature Review

Item response models are widely used for the analysis of psychometric data. Their origin can be traced back to mid-thirties and early forties (cf. Richardson, 1936; Tucker, 1946). A systematic development of item response theory from the classical point of view owes much to the pioneering work of Lord (1952, 1953a, 1953b), Rasch (1960, 1961) and their associates. Among the many noteworthy recent contributions in the same vein, we may cite Andersen (1970, 1972, 1973), Bock and Lieberman (1970), Mislevy and Bock (1984), Swaminathan and Gifford (1981), and Hambleton and Swaminathan (1985).

To motivate item response models, consider as a specific example, ability tests or attitude tests where each individual answers a battery of questions. Let X_{ij} denote the response of the i th individual to the j th question ($i = 1, \dots, n$; $j = 1, \dots, k$). Associated with the i th individual (or subject) is a subject parameter θ_i that expresses the capacity, ability or the attitude of that individual in a given context. However, the distributions of the X_{ij} will depend not only on θ_i but also on some parameter α_j where $-\alpha_j$ represents the difficulty level of question j . Item response analysis models the distribution of X_{ij} taking into account both θ_i and α_j ($i = 1, \dots, n$; $j = 1, \dots, k$).

For simplicity, throughout this dissertation we consider only the case when the X_{ij} are binary random variables. This is, for example, the situation when n examinees are

answering “True/False” questions. Alternately, the examinees may answer multiple choice questions where each answer is coded as “correct” or “incorrect”. Let $p_{ij} = P(X_{ij} = 1)$ ($i = 1, \dots, n$; $j = 1, \dots, k$), where 1 denotes a correct response. Item response theory for binary response is based on models which express the p_{ij} as functions of certain parameters.

This dissertation concentrates on one- and two-parameter item response models. A one-parameter item response model in its most general form is given by

$$p_{ij} = F(\theta_i + \alpha_j), \quad (1.1.1)$$

where F is a distribution function. This is referred to as a one-parameter item response model, since as a function of θ_i the right hand side of (1.1.1) has the form of a distribution function with location parameter $-\alpha_j$. The function F^{-1} (or sometimes F itself) is called a *link function*. The three most commonly used link functions are the logit, probit and log-log links. Logit and probit links are symmetric, while the log-log link is asymmetric. A logit link is based on the logistic distribution function, namely

$$F(x) = \frac{\exp(x)}{1 + \exp(x)} = (1 + \exp(-x))^{-1}.$$

The resulting one-parameter item response model is the celebrated Rasch model (Rasch, 1960, 1961). A probit link is based on the normal distribution function. A log-log link is based on the extreme value distribution function.

In the item response literature, the θ_i are referred to as “subject ability” parameters, while the α_j are referred to as “item” parameters. Despite the simplicity of one-parameter item response models and in particular, the overwhelming popularity of the Rasch model (see for example the collection of essays in Fischer and Molenaar, 1995), these models are often criticized on the ground that they assume all the questions to have equal power to discriminate between “good” and “poor” students, and therefore do not involve any discrimination parameters in addition to the item

parameters. This problem is alleviated by introducing two-parameter item response models given by

$$p_{ij} = F(\gamma_j \theta_i + b_j) \quad (1.1.2)$$

or equivalently by

$$p_{ij} = F(\gamma_j(\theta_i + \alpha_j)), \quad (b_j = \gamma_j \alpha_j). \quad (1.1.3)$$

The γ_j are referred to as “discrimination” parameters ranging between $(0, \infty)$ for each j . Models of this type are referred to as two-parameter item response models, since as a function of θ_i , p_{ij} has the form of a distribution function with location parameter $-\alpha_j$ and scale parameter $1/\gamma_j$.

One may be interested in inference for the θ_i or for the α_j or simultaneously for both the θ_i and the α_j in one-parameter item response models. Similarly for two-parameter item response models, inference may involve one or more of the three sets of parameters $\{\theta_i\}$, $\{\alpha_j\}$ and $\{\gamma_j\}$. Our dissertation will primarily concentrate on inference for the $\{\alpha_j\}$ in one-parameter models and for $\{\gamma_j, \alpha_j\}$ in two-parameter models.

We shall now briefly review some of the existing estimation procedures for one- and two-parameter item response models. Classical estimation for the $\{\alpha_j\}$ or $\{\alpha_j, \gamma_j\}$ uses maximum likelihood estimation method. A major difficulty that arises in this context is that the maximum likelihood estimators (MLE) of the item parameters are inconsistent. This is an example of the well known Neyman-Scott phenomenon, first noted by Neyman and Scott (1948). They proved in the balanced one-way normal ANOVA model that as the number of cell means grows to infinity, the MLE of the error variance is inconsistent. Andersen (1970, 1972, 1973), in a series of articles, has discussed very extensively frequentist inference for the Rasch model. One of his findings is that when $k = 2$, but $n \rightarrow \infty$, the MLE of α_1 (or α_2) is an inconsistent estimator. A proof of the inconsistency of the MLE for general k is recently given in Ghosh (1995).

Andersen (1970) recommends avoiding the problem of inconsistent MLE's for the Rasch model by finding the MLE's of the α_j ($j = 1, \dots, k$) conditional on sufficient statistics for the nuisance parameters θ_i ($i = 1, \dots, n$). This approach, however, does not work when separate sufficient statistics do not exist for the nuisance parameters. This is, for instance, evidenced in the probit model, $p_{ij} = \Phi(\theta_i + \alpha_j)$, Φ being the standard normal distribution function, where sufficient statistics do not exist for the nuisance parameters $(\theta_1, \dots, \theta_n)$. In fact, the only item response model that admits nontrivial sufficient statistics for the θ_i is the Rasch model. Bock and Lieberman (1970) have advocated instead the use of marginal maximum likelihood estimates of the α_j ($j = 1, \dots, k$) by assigning some distribution to $\theta_1, \dots, \theta_n$. Then integrating out with respect to $\theta_1, \dots, \theta_n$, one finds the joint marginal distribution of X_{ij} ($i = 1, \dots, n$; $j = 1, \dots, k$) which involve only $\alpha_1, \dots, \alpha_k$ and the parameters of the distributions of the θ_i . Based on this distribution, one finds the MLE's of the $\alpha_1, \dots, \alpha_k$, usually referred to as the marginal MLE's. Similarly for two-parameter models one can employ marginal maximum likelihood estimation technique for the item parameters as well as discrimination parameters. In this case one finds the marginal distribution of X_{ij} ($i = 1, \dots, n$; $j = 1, \dots, k$) involving the item parameters $\alpha_1, \dots, \alpha_k$ and $\gamma_1, \dots, \gamma_k$. Bock and Aitkin (1981) proposed an EM algorithm to obtain marginal maximum likelihood estimates for these parameters in the context of two-parameter models.

An alternative and attractive estimation procedure for item response models is the Bayesian procedure. Over the years, this approach has received considerable attention. However, even the Bayesian approach admits variation in actual use. First, in a regular Bayesian approach, inference is based on a completely specified prior distribution for all the unknown parameters. However, unless one uses fully noninformative priors, such procedures will lack robustness against misspecified priors. As an alternative, one may proceed by estimating some or all of the prior parameters

from the marginal distributions of the observations, and using the estimated prior to derive the estimated posterior. This estimated posterior is used for inferential purposes. This is the so-called empirical Bayes (EB) approach which has gained rapid popularity over the past decade.

As an alternative to the EB method, one uses a hierarchical Bayes (HB) procedure which assigns a prior distribution (often vague) to the unknown parameters rather than estimating them from the marginal distributions of the observations. The similarity between the two approaches is that they both recognize the uncertainty in estimating the prior parameters. Indeed, an EB method is often viewed as an approximation to a HB method. The two approaches usually yield similar point estimates of the parameters of interest. But the advantage of the HB method over the EB method is that unlike the latter, the former yields more reliable estimates of the variances of the parameters of interest by modeling the uncertainty of the prior parameters in the form of a distribution rather than obtaining point estimates of the prior parameters.

We now briefly summarize the existing Bayesian work on item response models. It is usually the item parameters which are of primary interest to the authors of Bayesian literature. But, sometimes there are interests on the subject parameters as well besides the item parameters. Tsutakawa and Johnson (1990) employed Bayesian methods to estimate the subject ability parameters. They considered a three-parameter logistic model (of which the two-parameter model is a special case) and eventually used an EB approach for ability estimation. Birnbaum (1969) and Owen (1975) have provided Bayesian estimation of ability parameters when the item parameters are known. Owen (1975) assumed a normal prior with a zero mean and a unit standard deviation for the ability parameters whereas Birnbaum (1969) assumed a logistic distribution for the same. Some authors have used Bayesian procedures for the joint estimation of both the ability and the item parameters. Although the Bayesian estimation of ability parameters can be traced back to the mid-sixties, the

procedures for joint estimation of both the ability and the item parameters are more of a recent nature. Swaminathan (1986), Swaminathan and Gifford (1981, 1982, 1985) have provided the joint Bayesian estimation of both the ability and the item parameters. They considered a two-parameter logistic model.

As mentioned earlier, our interest lies in Bayesian estimation of item parameters alone. There are many Bayesian articles available on the estimation of item parameters that treat the subject ability parameters as nuisance. Albert (1992) developed a regular Bayesian procedure for estimating item parameters using a two-parameter logistic model. He used a standard normal distribution as the prior for the subject ability parameters and a diffused prior for the item parameters. Kim et al. (1994) used a HB approach for item parameter estimation in the context of two-parameter logistic model. Swaminathan and Gifford (1982, 1985, 1986), Mislevy (1986) and Hambleton and Swaminathan (1985) have employed HB method for item parameter estimation for two-parameter models with the logit link. These authors used a standard normal distribution as the prior for the ability parameters, a first stage normal prior for the item difficulty parameters and an inverse chi-distribution for the item discrimination parameters. At the second stage they eventually used diffused priors for the item difficulty hyperparameters.

1.2 Topic of this Dissertation

The Bayesian methods that have been used so far in the literature are always link specific. The authors have used a specific link function, usually either logit or probit and developed their estimation procedure for that link. Our primary objective is to provide a unified Bayesian methodology for item response models that can be implemented with any of the three links considered, namely, logit, probit and log-log. A major concern in choosing priors is potential impropriety of the resulting posteriors.

Throughout this dissertation we discuss carefully the choice of priors leading to proper posteriors.

In Chapter 2 we discuss classical Bayesian methodology for one-parameter item response models and provide a general algorithm to implement it in any given situation. We also provide conditions under which marginal ML estimators are consistent. Also in this chapter we devote a section to approximation of improper priors by proper ones.

In Chapter 3 we deal with hierarchical Bayesian methods for item response models. We develop a hierarchical approach for one-parameter models and discuss the important issues of propriety of posteriors and implementation.

In Chapter 4 we discuss classical Bayesian methodology for two-parameter models and develop a unified theory to implement them in a given context. In all these chapters we illustrate our methodology with the help of real life data on a mathematics placement test.

CHAPTER 2

A UNIFIED BAYESIAN ANALYSIS OF ITEM RESPONSE MODELS FOR BINARY DATA

2.1 Introduction

As discussed in the introduction, there is considerable non-Bayesian literature for one-parameter item response models. Possibly, the biggest attention has been paid to the Rasch model which has applications well beyond the psychology and sociology literature. The probit model has also been used in many different contexts.

Much of the Bayesian literature on this topic is devoted either to partial Bayes or empirical Bayes estimation procedure where some prior distribution has been assigned to the subject and/or the item parameters, and these parameters are either simultaneously estimated by their posterior modes, or the item parameters are estimated by their marginal model after integrating out the subject parameters. However, to this date, a complete unified Bayesian analysis seems to be lacking for these models. The present chapter makes an attempt in this direction.

The outline of the remaining sections in this chapter is as follows. It is shown that Laplace's prior necessarily leads to a nonidentifiable improper posterior. A necessary and sufficient condition for propriety of posteriors with nonidentifiable likelihood is provided. Next in this section, we consider flat priors for item parameters, and proper priors for subject parameters. Sufficient conditions are given under which these priors lead to proper posteriors. Implementation of the Bayes procedure via Markov Chain Monte Carlo integration technique is given in Section 2.3. The Bayesian technique is

illustrated with a placement examination dataset in Section 2.4 for the logit, probit and log-log links. Section 2.5 considers a specific example of matched pairs data, and compares and contrasts the Bayesian technique with the marginal and conditional likelihoods. Section 2.6 discusses uniform approximation of improper priors by proper priors. This extends the work of Mukhopadhyay and Dasgupta (1992) and Mukhopadhyay and Ghosh (1994) for location-scale models. Finally, Section 2.7 gives a formal proof of the consistency of the marginal maximum likelihood estimators for fairly general link functions.

2.2 Choice Of Priors

For the one-parameter item response model, writing $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)$, $\mathbf{x} = (x_{11}, \dots, x_{1k}, \dots, x_{n1}, \dots, x_{nk})$, and $\bar{F} = 1 - F$, the likelihood function is given by

$$L(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{x}) = \prod_{i=1}^n \prod_{j=1}^k \left[F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j) \right], \quad (2.2.1)$$

We now consider different choices of priors for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$. One straightforward approach is to use a subjective proper prior distribution $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha})$ for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, and find the posterior distribution for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ as

$$\pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{x}) \propto L(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{x}) \pi(\boldsymbol{\theta}, \boldsymbol{\alpha}). \quad (2.2.2)$$

However, often the prior information is vague, in which case it is useful to consider noninformative priors.

The simplest noninformative prior, due to Laplace, is $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}) \propto 1$. However, this choice of priors leads to improper posteriors for $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$. This fact is a consequence of the following lemma. This lemma is possibly known to many Bayesians, and seems to be implicit in Dawid (1979) and O'Hagan (1994) (p 72 and p 158). But the following explicit formulation seems worthwhile for future use.

Lemma 1.

Suppose the parameter vector is (ϕ_1, ϕ_2) where either or both of ϕ_1 and ϕ_2 may be vector valued. Suppose $\mathbf{X}$ has a nonidentifiable pdf $f(\mathbf{x}|\phi_1, \phi_2) = f(\mathbf{x}|\phi_1)$. Consider the prior $\pi(\phi_1, \phi_2)$ for (ϕ_1, ϕ_2) . Then the posterior $\pi(\phi_1, \phi_2|\mathbf{x})$ is proper if and only if $\pi(\phi_1|\mathbf{x})$ and $\pi(\phi_2|\phi_1)$ are both proper.

Proof.

$$\begin{aligned}\pi(\phi_1, \phi_2|\mathbf{x}) &\propto f(\mathbf{x}|\phi_1, \phi_2)\pi(\phi_1, \phi_2) \\ &= f(\mathbf{x}|\phi_1)\pi(\phi_1)\pi(\phi_2|\phi_1) \\ &\propto \pi(\phi_1|\mathbf{x})\pi(\phi_2|\phi_1).\end{aligned}$$

This proves the lemma.

Using the above lemma, it is possible to prove the following theorem showing that Laplace's prior necessarily leads to an improper posterior in the present context.

Theorem 1.

Consider the likelihood function given in (2.2.1), and the prior $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}) \propto 1$. Then the posterior $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathbf{x})$ is improper, that is

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathbf{x}) d\boldsymbol{\theta} d\boldsymbol{\alpha} = \infty. \quad (2.2.3)$$

Proof.

Make the one to one linear transformation

$$\begin{aligned}\eta_i &= \theta_i + \alpha_k \quad i = 1, \dots, n, \\ \xi_j &= \alpha_j - \alpha_k \quad j = 1, \dots, k-1.\end{aligned}$$

Write $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{k-1})$. Then $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ is one to one with $(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k)$ and the likelihood function given in (2.2.1) can be rewritten as

$$L(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k) = \prod_{i=1}^n \prod_{j=1}^{k-1} \left[F^{x_{ij}}(\eta_i + \xi_j) \bar{F}^{1-x_{ij}}(\eta_i + \xi_j) \right] \prod_{i=1}^n \left[F^{x_{ik}}(\eta_i) \bar{F}^{1-x_{ik}}(\eta_i) \right] \quad (2.2.4)$$

Since the Jacobian of the transformation from $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ to $(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k)$ is a constant free of any parameter, $\pi(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k) \propto 1$. This implies $\pi(\alpha_k | \boldsymbol{\eta}, \boldsymbol{\xi}) \propto 1$ which is improper. Now apply Lemma 1 with $\phi_2 = \alpha_k$ and $\phi_1 = (\boldsymbol{\eta}, \boldsymbol{\xi})$ to conclude that $\pi(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k | \mathbf{x})$ is improper. Hence,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{x}) d\boldsymbol{\theta} d\boldsymbol{\alpha} \propto \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k | \mathbf{x}) d\boldsymbol{\eta} d\boldsymbol{\xi} d\alpha_k = \infty.$$

Remark 1. It is hinted in Swaminathan and Gifford (1985, p 353) that using flat priors for both $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, the Bayesian analysis will be equivalent to likelihood based analysis. But the resulting posterior distribution being improper, it will make very little sense to talk about posterior mean, Bayesian credible sets, posterior quantiles and other posterior quantities.

Remark 2. The nonidentifiability of the likelihood becomes apparent with the reparametrization given in (2.2.4). A similar nonidentifiability occurs for more complex item response models as well. Also the result as stated in Lemma 1 should be of use in similar other contexts.

Theorem 1 raises the question whether the subset $(\boldsymbol{\eta}, \boldsymbol{\xi})$ of $(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k)$ has a proper posterior, since the nonidentifiability problem then disappears. However, the answer continues to be negative when either *for at least one i* , the x_{ij} ($j = 1, \dots, k$) are all zeros or 1's, or *for at least one j* , the x_{ij} ($i = 1, \dots, n$) are all zeros or all 1's. To see this consider the likelihood as given in (2.2.4) as $L(\boldsymbol{\eta}, \boldsymbol{\xi})$. Also, let

$$t_i = \sum_{j=1}^k x_{ij} \quad i = 1, \dots, n,$$

$$y_j = \sum_{i=1}^n x_{ij} \quad j = 1, \dots, k.$$

Then, we prove the following result.

Theorem 2.

Suppose for at least one i , $t_i = 0$ or k . Then with the prior $\pi(\boldsymbol{\eta}, \boldsymbol{\xi}) \propto 1$, $\pi(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{x})$ is

improper. Also, if $y_j = 0$ or n for at least one j , $\pi(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{x})$ continues to be improper under the same prior.

Proof.

Suppose $t_m = 0$, that is $x_{m1} = \dots = x_{mk} = 0$. Now,

$$\int_{-\infty}^{\infty} \prod_{j=1}^{k-1} \bar{F}(\eta_m + \xi_j) \bar{F}(\eta_m) d\eta_m \geq \int_{-\infty}^0 \prod_{j=1}^{k-1} \bar{F}(\xi_j) \bar{F}(0) d\eta_m = +\infty. \quad (2.2.5)$$

Similarly, if $t_m = k$, that is $x_{m1} = \dots = x_{mk} = 1$,

$$\int_{-\infty}^{\infty} \prod_{j=1}^{k-1} F(\eta_m + \xi_j) F(\eta_m) d\eta_m \geq \int_0^{\infty} \prod_{j=1}^{k-1} F(\xi_j) F(0) d\eta_m = +\infty. \quad (2.2.6)$$

Also, if $y_l = 0$, that is $x_{1l} = \dots = x_{nl} = 0$, one gets

$$\begin{aligned} \int_{-\infty}^{\infty} \prod_{i=1}^n \bar{F}(\eta_i + \xi_l) d\xi_l &\geq \int_{-\infty}^0 \prod_{i=1}^n \bar{F}(\eta_i + \xi_l) d\xi_l \\ &\geq \prod_{i=1}^n \bar{F}(\eta_i) \int_{-\infty}^0 d\xi_l = +\infty. \end{aligned} \quad (2.2.7)$$

Finally, if $y_l = n$, that is $x_{1l} = \dots = x_{nl} = 1$,

$$\begin{aligned} \int_{-\infty}^{\infty} \prod_{i=1}^n F(\eta_i + \xi_l) d\xi_l &\geq \int_0^{\infty} \prod_{i=1}^n F(\eta_i + \xi_l) d\xi_l \\ &\geq \prod_{i=1}^n F(\eta_i) \int_0^{\infty} d\xi_l = +\infty. \end{aligned} \quad (2.2.8)$$

The theorem follows now from (2.2.5) - (2.2.8).

It appears, however, that once the boundary values are excluded for the t_i and the y_j , the posterior $\pi(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{x})$ is proper. This result is proved for $k = 2$ when the link function is logit, probit or log-log. It is conjectured that the result is true for an arbitrary k , but we have not found a formal proof as yet.

Theorem 3.

Suppose $k = 2$, $\pi(\boldsymbol{\eta}, \boldsymbol{\xi}) \propto 1$, and $0 < y_j < n$ and $t_i = 1$ all i and j . Then the posterior $\pi(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{x})$ is proper when (i) $F(x) = \exp(x)/[1 + \exp(x)]$, (ii) $F(x) = \Phi(x)$, the standard normal distribution function, or (iii) $F(x) = \exp(-\exp(-x))$.

Proof.

First consider the logit link. Since $t_i = 1$ for all $i = 1, \dots, n$, the posterior is,

$$\pi(\boldsymbol{\eta}, \xi_1 | \mathbf{x}) = \frac{\exp(\sum_{i=1}^n \eta_i + \xi_1 y_1)}{\prod_{i=1}^n [\{1 + \exp(\eta_i + \xi_1)\} \{1 + \exp(\eta_i)\}]} \quad (2.2.9)$$

We want to show that

$$I = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi(\boldsymbol{\eta}, \xi_1 | \mathbf{x}) d\boldsymbol{\eta} d\xi_1 < \infty. \quad (2.2.10)$$

Substituting $z = \exp(\eta_i)$,

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{\exp(\eta_i)}{[\{1 + \exp(\eta_i + \xi_1)\} \{1 + \exp(\eta_i)\}]} d\eta_i \\ &= \int_0^{\infty} [\{1 + z \exp(\xi_1)\} (1 + z)]^{-1} dz \\ &= \xi_1 [\exp(\xi_1) - 1]^{-1}. \end{aligned} \quad (2.2.11)$$

Then, substituting $u = \exp(\xi_1)$,

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \exp(\xi_1 y_1) \xi_1^n [\exp(\xi_1) - 1]^{-n} d\xi_1 \\ &= \int_0^{\infty} (\log u)^n (u - 1)^{-n} u^{y_1 - 1} du \\ &= \int_0^2 (\log u)^n (u - 1)^{-n} u^{y_1 - 1} du + \int_2^{\infty} (\log u)^n (u - 1)^{-n} u^{y_1 - 1} du \\ &\leq \int_0^2 u^{y_1 - 1} du + \int_2^{\infty} (\log u)^n u^{y_1 - 1} \left(\frac{1}{2}u\right)^{-n} du \\ &= 2^{y_1} y_1^{-1} + 2^n \int_2^{\infty} (\log u)^n u^{-2} du \\ &= 2^{y_1} y_1^{-1} + 2^n \int_{\log 2}^{\infty} e^{-v} v^n dv < \infty. \end{aligned} \quad (2.2.12)$$

Next consider probit link with $k = 2$. The condition of the theorem says all the data points are either (1,0) or (0,1). For this link the posterior is

$$\pi(\boldsymbol{\eta}, \xi_1 | \mathbf{x}) = \prod_{i=1}^n \Phi^{x_{i1}}(\eta_i + \xi_1) \bar{\Phi}^{1-x_{i1}}(\eta_i + \xi_1) \Phi^{x_{i2}}(\eta_i) \bar{\Phi}^{1-x_{i2}}(\eta_i), \quad (2.2.13)$$

where $\bar{\Phi} = 1 - \Phi$. We want to show that

$$I = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi(\boldsymbol{\eta}, \xi_1 | \mathbf{x}) d\boldsymbol{\eta} d\xi_1 < \infty.$$

Let n_{10} =number of subjects for which $x_{i1} = 1, x_{i2} = 0$ and n_{01} =number of subjects for which $x_{i1} = 0, x_{i2} = 1$. Then,

$$I = \int_{-\infty}^{\infty} \left[\int_{-\infty}^{\infty} \Phi(\eta + \xi_1) \bar{\Phi}(\eta) d\eta \right]^{n_{10}} \left[\int_{-\infty}^{\infty} \bar{\Phi}(\eta + \xi_1) \Phi(\eta) d\eta \right]^{n_{01}} d\xi_1. \quad (2.2.14)$$

Now, writing $\phi(u) = \Phi'(u)$, and integrating by parts,

$$\begin{aligned} & \int_{-\infty}^{\infty} \Phi(\eta + \xi_1) \bar{\Phi}(\eta) d\eta \\ &= [\Phi(\eta + \xi_1) \bar{\Phi}(\eta) \eta]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} \eta [\phi(\eta + \xi_1) \bar{\Phi}(\eta) - \Phi(\eta + \xi_1) \phi(\eta)] d\eta \\ &= \int_{-\infty}^{\infty} \Phi(\eta + \xi_1) \eta \phi(\eta) d\eta - \int_{-\infty}^{\infty} \bar{\Phi}(\eta) \eta \phi(\eta + \xi_1) d\eta; \end{aligned} \quad (2.2.15)$$

Integrating by parts,

$$\begin{aligned} \int_{-\infty}^{\infty} \Phi(\eta + \xi_1) \eta \phi(\eta) d\eta &= - \int_{-\infty}^{\infty} \Phi(\eta + \xi_1) \phi'(\eta) d\eta \\ &= \int_{-\infty}^{\infty} \phi(\eta + \xi_1) \phi(\eta) d\eta \\ &= \int_{-\infty}^{\infty} (2\pi^{-1} \exp(-\frac{1}{2}[(\eta + \xi_1)^2 + \eta^2])) d\eta \\ &= (2\pi^{1/2})^{-1} \exp(-\frac{1}{4}\xi_1^2). \end{aligned} \quad (2.2.16)$$

Next, integrating by parts, and using the standard convolution formula,

$$\begin{aligned}
& - \int_{-\infty}^{\infty} \bar{\Phi}(\eta) \eta \phi(\eta + \xi_1) d\eta \\
&= \xi_1 \int_{-\infty}^{\infty} \bar{\Phi}(\eta - \xi_1) \phi(\eta) d\eta - \int_{-\infty}^{\infty} \bar{\Phi}(\eta - \xi_1) \eta \phi(\eta) d\eta \\
&= \xi_1 \int_{-\infty}^{\infty} \bar{\Phi}(\xi_1 - \eta) \phi(\eta) d\eta + \int_{-\infty}^{\infty} \bar{\Phi}(\eta - \xi_1) \phi'(\eta) d\eta \\
&= \xi_1 P(N(0, 2) \leq \xi_1) + \int_{-\infty}^{\infty} \phi(\eta) \phi(\eta - \xi_1) d\eta \\
&= \xi_1 \Phi(\xi_1/\sqrt{2}) + (2\pi^{1/2})^{-1} \exp(-\frac{1}{4}\xi_1^2). \tag{2.2.17}
\end{aligned}$$

Hence,

$$\int_{-\infty}^{\infty} \Phi(\eta + \xi_1) \bar{\Phi}(\eta) d\eta = \xi_1 \Phi(\xi_1/\sqrt{2}) + \pi^{-\frac{1}{2}} \exp(-\frac{1}{4}\xi_1^2). \tag{2.2.18}$$

Similar calculations give

$$\int_{-\infty}^{\infty} \bar{\Phi}(\eta + \xi_1) \Phi(\eta) d\eta = -\xi_1 \Phi(\xi_1/\sqrt{2}) + \pi^{-\frac{1}{2}} \exp(-\frac{1}{4}\xi_1^2). \tag{2.2.19}$$

Hence,

$$\begin{aligned}
I &= \int_{-\infty}^{\infty} \left[\pi^{-\frac{1}{2}} \exp\left(-\frac{1}{4}\xi_1^2\right) + \xi_1 \Phi(\xi_1/\sqrt{2}) \right]^{n_{10}} \times \\
&\quad \left[\pi^{-\frac{1}{2}} \exp\left(-\frac{1}{4}\xi_1^2\right) - \xi_1 \Phi(-\xi_1/\sqrt{2}) \right]^{n_{01}} d\xi_1 \\
&\leq \int_0^{\infty} \left(\pi^{-\frac{1}{2}} \exp(-\frac{\xi_1^2}{4}) \right)^{n_{01}} [\pi^{-\frac{1}{2}} + \xi_1]^{n_{10}} d\xi_1 \\
&\quad + \int_0^{\infty} \left(\pi^{-\frac{1}{2}} \exp(-\frac{\xi_1^2}{4}) \right)^{n_{10}} [\pi^{-\frac{1}{2}} + \xi_1]^{n_{01}} d\xi_1 < \infty. \tag{2.2.20}
\end{aligned}$$

Finally consider the log-log link. Then, the posterior is

$$\begin{aligned}
\pi(\boldsymbol{\eta}, \xi_1 | \mathbf{x}) &= \left[\prod_{i: x_{i1}=1, x_{i2}=0} \exp(-\exp(-\eta_i - \xi_1))(1 - \exp(-\exp(-\eta_i))) \right] \\
&\times \left[\prod_{i: x_{i1}=0, x_{i2}=1} \exp(-\exp(-\eta_i))(1 - \exp(-\exp(-\eta_i - \xi_1))) \right].
\end{aligned} \tag{2.2.21}$$

We want to show that

$$I = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \pi(\boldsymbol{\eta}, \xi_1 | \mathbf{x}) d\boldsymbol{\eta} d\xi_1 < \infty.$$

Now, writing $\exp(-\eta_i) = z$ and $\exp(-\xi_1) = a$, and integrating by parts

$$\begin{aligned}
&\int_{-\infty}^{\infty} \exp(-\exp(-\eta_i - \xi_1))(1 - \exp(-\exp(-\eta_i))) d\eta_i \\
&= \int_0^{\infty} \exp(-z \exp(-\xi_1))(1 - \exp(-z)) z^{-1} dz \\
&= \int_0^{\infty} \log z [-a \exp(-za) + (a+1) \exp(-z(a+1))] dz \\
&= a \int_0^{\infty} \exp(-za) [\log(za) - \log a] dz \\
&\quad - (a+1) \int_0^{\infty} \exp(-z(a+1)) [\log(z(a+1)) - \log(a+1)] dz \\
&= \int_0^{\infty} \exp(-y) \log y dy - \log a \int_0^{\infty} \exp(-y) dy \\
&\quad - \int_0^{\infty} \exp(-y) \log y dy + \log(a+1) \int_0^{\infty} \exp(-y) dy \\
&= \log(1 + a^{-1}) = \log[1 + \exp(\xi_1)].
\end{aligned} \tag{2.2.22}$$

Similarly,

$$\int_{-\infty}^{\infty} \exp(-\exp(-\eta_i))(1 - \exp(-\exp(-\eta_i - \xi_1))) d\eta_i = \log[1 + \exp(-\xi_1)]. \tag{2.2.23}$$

Thus,

$$\begin{aligned}
I &= \int_{-\infty}^{\infty} [\log(1 + \exp(\xi_1))]^{n_{10}} [\log(1 + \exp(-\xi_1))]^{n_{01}} d\xi_1 \\
&\leq \int_0^{\infty} \exp(-\xi_1 n_{01}) [\log(1 + \exp(\xi_1))]^{n_{10}} d\xi_1 \\
&\quad + \int_{-\infty}^0 \exp(\xi_1 n_{10}) [\log(1 + \exp(-\xi_1))]^{n_{01}} d\xi_1 < \infty.
\end{aligned} \tag{2.2.24}$$

This completes the proof.

As an immediate consequence of the above findings, it follows that the joint posterior of $(\xi_1, \dots, \xi_{k-1}) = (\alpha_1 - \alpha_k, \dots, \alpha_{k-1} - \alpha_k)$ given $\mathbf{x}$ may be improper when $t_i = 0$ or k for at least one i , or $y_j = 0$ or n for at least one j . This is in contrast to the conclusions of the usual one way normal ANOVA model considered for example, in Sahu and Gelfand (1995). To see this suppose $Y_1, \dots, Y_k$ are independent $N(\mu + \alpha_j, 1)$, ($j = 1, \dots, k$). If one assigns Laplace's prior $\pi(\mu, \alpha_1, \dots, \alpha_k) \propto 1$, then the joint posterior of $(\mu, \alpha_1, \dots, \alpha_k)$ is improper, once again as an immediate consequence of Lemma 1. Yet the joint posterior of $(\mu + \alpha_1, \dots, \mu + \alpha_k)$ is a product of independent normals. This implies that the joint posterior of $(\alpha_1 - \alpha_k, \dots, \alpha_{k-1} - \alpha_k) = ((\mu + \alpha_1) - (\mu + \alpha_k), \dots, (\mu + \alpha_{k-1}) - (\mu + \alpha_k))$ is also proper. This difference seems mainly due to the different likelihood and link structures for the normal and item response models. Sahu and Gelfand (1995) provide a detailed discussion of how estimability of parameters can be linked with the propriety or otherwise of posteriors for the normal model. A similar phenomenon seems unavailable here due to the inherent nonlinear structure of the model.

Clearly from Lemma 1 and the reparametrizations introduced in (2.2.4), one gets a necessary and sufficient condition for the posterior of $(\boldsymbol{\eta}, \boldsymbol{\xi}, \boldsymbol{\alpha}_k)$ given $\mathbf{x}$ be proper. This is equivalent to the propriety of the posterior of $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ given $\mathbf{x}$. However, verification of the conditions of Lemma 1 after the necessary reparametrization need not always be easy. We find it more convenient to verify the propriety of posteriors

of (θ, α) by direct calculations for the logit, probit and log-log link. Specifically consider the prior $\pi(\theta, \alpha) \propto g(\theta)$ when g is a proper pdf. The following theorem provides specific conditions under which $\pi(\theta, \alpha | \mathbf{x})$ is proper.

Theorem 4.

Assume the conditions (I) $g(\theta)$ has finite moment generating function (mgf) and (II) $1 \leq y_j \leq n-1$ for all $j = 1, \dots, k$. Then for the noninformative prior on α , $\pi(\theta, \alpha | \mathbf{x})$ is proper for the link functions (i)-(iii) of Theorem 3.

Proof.

(i) First consider the logit link. In this case, the likelihood function is given by

$$L(\theta, \alpha | \mathbf{x}) = \frac{\exp\left(\sum_{i=1}^n \theta_i t_i + \sum_{j=1}^k \alpha_j y_j\right)}{\prod_{i=1}^n \prod_{j=1}^k (1 + \exp(\theta_i + \alpha_j))}. \quad (2.2.25)$$

We want to show

$$I = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L(\theta, \alpha | \mathbf{x}) g(\theta) \prod_{i=1}^n d\theta_i \prod_{j=1}^k d\alpha_j < \infty \quad (2.2.26)$$

provided $g(\theta)$ has finite mgf. First we get the inequality,

$$\begin{aligned} & \int_{-\infty}^{\infty} \frac{\exp(\alpha y_j)}{\prod_{i=1}^n (1 + \exp(\theta_i + \alpha))} d\alpha \\ &= \int_{-\infty}^0 \frac{\exp(\alpha y_j)}{\prod_{i=1}^n (1 + \exp(\theta_i + \alpha))} d\alpha + \int_0^{\infty} \frac{\exp(\alpha y_j)}{\prod_{i=1}^n (1 + \exp(\theta_i + \alpha))} d\alpha \\ &\leq \int_{-\infty}^0 \exp(\alpha y_j) d\alpha + \exp\left(-\sum_1^n \theta_i\right) \int_0^{\infty} \exp(-(n - y_j)\alpha) d\alpha \\ &\leq y_j^{-1} + \exp\left(-\sum_1^n \theta_i\right) (n - y_j)^{-1} \leq 1 + \exp\left(-\sum_1^n \theta_i\right). \end{aligned} \quad (2.2.27)$$

Hence,

$$I \leq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left(\sum_{i=1}^n \theta_i t_i\right) [1 + \exp\left(-\sum_1^n \theta_i\right)]^k g(\theta) \prod_{i=1}^n d\theta_i < \infty, \quad (2.2.28)$$

due to the finiteness of the mgf of g . This proves (i).

(ii) In this case

$$L(\theta, \alpha | \mathbf{x}) = \prod_{i=1}^n \prod_{j=1}^k \left[\Phi^{x_{ij}}(\theta_i + \alpha_j) (1 - \Phi(\theta_i + \alpha_j))^{1-x_{ij}} \right]. \quad (2.2.29)$$

First, using the inequality $\frac{1}{2}[\exp(t) + \exp(-t)] \leq \exp(t^2/2)$ for all real t , note that for $\alpha_j > 0$,

$$\begin{aligned} 1 - \Phi(\theta_i + \alpha_j) &= \int_{\theta_i + \alpha_j}^{\infty} (2\pi)^{-1/2} \exp(-t^2/2) dt \\ &\leq (2\pi)^{-1/2} \int_{\theta_i + \alpha_j}^{\infty} 2(e^t + e^{-t})^{-1} dt \\ &\leq (2/\pi)^{1/2} \int_{\theta_i + \alpha_j}^{\infty} e^{-t} dt \\ &= (2/\pi)^{1/2} \exp(-\theta_i - \alpha_j). \end{aligned} \quad (2.2.30)$$

Similarly, for $\alpha_j < 0$,

$$\begin{aligned} \Phi(\theta_i + \alpha_j) &\leq (2/\pi)^{1/2} \int_{-\infty}^{\theta_i + \alpha_j} e^t dt \\ &\leq (2/\pi)^{1/2} \exp(\theta_i + \alpha_j). \end{aligned} \quad (2.2.31)$$

Hence,

$$\begin{aligned} I &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} L(\theta, \alpha | \mathbf{x}) g(\theta) \prod_{i=1}^n d\theta_i \prod_{j=1}^k d\alpha_j \\ &\leq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\theta) \prod_{j=1}^k \left[\int_{-\infty}^0 \{(2/\pi)^{1/2} \exp(\theta_i + \alpha_j)\}^{x_{ij}} d\alpha_j \right. \\ &\quad \left. + \int_0^{\infty} \{(2/\pi)^{1/2} \exp(-\theta_i - \alpha_j)\}^{1-x_{ij}} d\alpha_j \right] d\theta_1 \cdots d\theta_n \\ &\leq \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(\theta) \prod_{j=1}^k [(2/\pi)^{y_j/2} \exp(\sum_{i=1}^n x_{ij} \theta_i) \int_{-\infty}^0 \exp(\alpha y_j) d\alpha \end{aligned}$$

$$\begin{aligned}
& + (2/\pi)^{n-y_j/2} \exp\left(-\sum_{i=1}^n (1-x_{ij})\theta_i\right) \int_0^\infty \exp(-(n-y_j)\alpha) d\alpha \prod_{i=1}^n d\theta_i \\
& \leq \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty g(\theta) (2/\pi)^{k/2} \prod_{j=1}^k [\exp(\sum_{i=1}^n x_{ij}\theta_i) y_j^{-1} \\
& \quad + \exp(-\sum_{i=1}^n (1-x_{ij})\theta_i) (n-y_j)^{-1}] \prod_{i=1}^n d\theta_i \\
& \leq (2/\pi)^{k/2} \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty g(\theta) \prod_{j=1}^k [\exp(\sum_{i=1}^n x_{ij}\theta_i) + \exp(-\sum_{i=1}^n (1-x_{ij})\theta_i)] \prod_{i=1}^n d\theta_i \\
& = (2/\pi)^{k/2} \int_{-\infty}^\infty \cdots \int_{-\infty}^\infty g(\theta) \exp(\sum_1^n t_i \theta_i) (1 + \exp(-\sum_1^n \theta_i))^k \prod_{i=1}^n d\theta_i \\
& < \infty.
\end{aligned} \tag{2.2.32}$$

due to the finiteness of the mgf of $g(\theta)$. This proves (ii).

(iii) In this case

$$\begin{aligned}
L(\theta, \alpha | \mathbf{x}) &= \exp\left(-\sum_{i=1}^n \sum_{j=1}^k x_{ij} \exp(-\theta_i - \alpha_j)\right) \\
&\times \prod_{i=1}^n \prod_{j=1}^k [1 - \exp(-\exp(-\theta_i - \alpha_j))].
\end{aligned} \tag{2.2.33}$$

We want to show that the integral

$$I = \int_{\mathbf{R}^n} \int_{\mathbf{R}^k} g(\theta) L(\theta, \alpha | \mathbf{x}) d\alpha d\theta < \infty. \tag{2.2.34}$$

But

$$\begin{aligned}
I &\leq \int_{\mathbf{R}^n} g(\theta) \prod_{j=1}^k \left[\int_{-\infty}^0 \exp\left(-\sum_i x_{ij} \exp(-\theta_i - \alpha_j)\right) d\alpha_j \right. \\
&\quad \left. + \int_0^\infty \prod_{i=1}^n [1 - \exp(-\exp(-\theta_i - \alpha_j))]^{1-x_{ij}} d\alpha_j \right] d\theta.
\end{aligned} \tag{2.2.35}$$

Now using $\exp(-u) \leq u^{-1}$ for $u > 0$,

$$\begin{aligned}
& \int_{-\infty}^0 \exp\left(-\sum_i^n x_{ij} \exp(-\theta_i - \alpha_j)\right) d\alpha_j \\
&= \int_{-\infty}^0 \exp\left(-\sum_i^n x_{ij} \exp(-\theta_i) \exp(-\alpha_j)\right) d\alpha_j \\
&\leq \int_{-\infty}^0 \left(\sum_{i=1}^n x_{ij} \exp(-\theta_i)\right)^{-1} \exp(\alpha_j) d\alpha_j \\
&= \left(\sum_{i=1}^n x_{ij} \exp(-\theta_i)\right)^{-1} \leq [\max_{1 \leq i \leq n} \exp(|\theta_i|)] y_j^{-1} \leq \sum_{i=1}^n \exp(|\theta_i|). \quad (2.2.36)
\end{aligned}$$

Also, using $1 - \exp(-u) \leq u$ for all $u > 0$,

$$\begin{aligned}
& \int_0^\infty \prod_{i=1}^n [1 - \exp(-\exp(-\theta_i - \alpha_j))]^{1-x_{ij}} d\alpha_j \\
&\leq \int_0^\infty \prod_{i=1}^n \{\exp(-\theta_i - \alpha_j)\}^{1-x_{ij}} d\alpha_j \\
&= \exp\left(-\sum_1^n \theta_i (1 - x_{ij})\right) \int_0^\infty \exp(-\alpha_j (n - y_j)) d\alpha_j \\
&= \exp\left(-\sum_1^n \theta_i (1 - x_{ij})\right) (n - y_j)^{-1} \leq \exp\left(\sum_{i=1}^n |\theta_i|\right). \quad (2.2.37)
\end{aligned}$$

Hence, from (2.2.35) - (2.2.37),

$$I \leq \int_{\mathbf{R}^n} g(\boldsymbol{\theta}) [2 \exp(\sum_1^n |\theta_i|)]^k d\boldsymbol{\theta} < \infty, \quad (2.2.38)$$

due to the finiteness of the mgf of $g(\boldsymbol{\theta})$. This proves (iii), and completes the proof of Theorem 4.

Before concluding this section, we establish an invariance property of the posterior means of the elementary contrasts $\alpha_j - \alpha_m$ ($1 \leq j \neq m \leq k$) for general link functions and the location-scale family of priors $g_{\mu, \sigma}(\boldsymbol{\theta}) = \sigma^{-n} g((\theta_1 - \mu)/\sigma, \dots, (\theta_n - \mu)/\sigma)$ for

θ and independent flat prior for α . The following theorem shows that the Bayes estimate of $\alpha_j - \alpha_m$ is location invariant, that is it does not depend on the choice of μ . Thus as in Section 4 of this paper, if one is interested in inference about the elementary contrasts, for a location-scale prior, μ can be taken as 0 without loss of generality.

Theorem 5.

Consider the likelihood given in (2.2.1) and the prior

$$\pi(\theta, \alpha) \propto \sigma^{-n} g((\theta_1 - \mu)/\sigma, \dots, (\theta_n - \mu)/\sigma). \quad (2.2.39)$$

Then the posterior $\pi_{\mu, \sigma}(\alpha | \mathbf{x})$ of α equals $\pi_{0, \sigma}(\gamma | \mathbf{x})$, where $\gamma = (\gamma_1, \dots, \gamma_k)$ and $\gamma_j = \mu + \alpha_j$ ($j = 1, \dots, k$).

Proof.

Let $z_i = \theta_i - \mu$ ($i = 1, \dots, n$), $\mathbf{z} = (z_1, \dots, z_n)$. Then

$$\begin{aligned} \pi_{\mu, \sigma}(\alpha | \mathbf{x}) &= \frac{\int_{R^n} L(\theta, \alpha | \mathbf{x}) \pi(\theta, \alpha) d\theta}{\int_{R^k} \int_{R^n} L(\theta, \alpha | \mathbf{x}) \pi(\theta, \alpha) d\theta d\alpha} \\ &= \int_{R^n} \prod_{i=1}^n \prod_{j=1}^k \left\{ F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j) \right\} \frac{1}{\sigma^n} g\left(\frac{\theta_1 - \mu}{\sigma}, \dots, \frac{\theta_n - \mu}{\sigma}\right) d\theta \\ &\div \int_{R^k} \int_{R^n} \prod_{i=1}^n \prod_{j=1}^k \left\{ F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j) \right\} \frac{1}{\sigma^n} g\left(\frac{\theta_1 - \mu}{\sigma}, \dots, \frac{\theta_n - \mu}{\sigma}\right) d\theta d\alpha \\ &= \int_{R^n} \prod_{i=1}^n \prod_{j=1}^k \left\{ F^{x_{ij}}(z_i + \gamma_j) \bar{F}^{1-x_{ij}}(z_i + \gamma_j) \right\} g\left(\frac{z_1}{\sigma}, \dots, \frac{z_n}{\sigma}\right) dz \\ &\div \int_{R^k} \int_{R^n} \prod_{i=1}^n \prod_{j=1}^k \left\{ F^{x_{ij}}(z_i + \gamma_j) \bar{F}^{1-x_{ij}}(z_i + \gamma_j) \right\} g\left(\frac{z_1}{\sigma}, \dots, \frac{z_n}{\sigma}\right) dz d\gamma \\ &= \pi_{0, \sigma}(\gamma | \mathbf{x}). \end{aligned}$$

Remark 3. Since $\alpha_j - \alpha_m = \gamma_j - \gamma_m$ ($1 \leq j \neq m \leq k$), the joint posterior of these contrasts does not depend on μ .

2.3 Implementation Of Bayes Procedures

Consider the general link $p_{ij} = F(\theta_i + \alpha_j)$, $i = 1, \dots, n$, $j = 1, \dots, k$, and the prior $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}) \propto \prod_{i=1}^n g_1(\theta_i) \prod_{j=1}^k g_2(\alpha_j)$. Both g_1 and g_2 can be improper as long as the posterior $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{x})$ remains proper. However, due to nonconjugacy of the prior, the posterior is analytically intractable, and can be found only via numerical integration. Also, direct numerical integration seems infeasible in this case because of the high dimensionality of the problem.

Fortunately, the integration task has become easier due to the advent of the sophisticated Markov Chain Monte Carlo (MC²) numerical integration techniques. We shall use, in particular, Gibbs sampling for integration purposes. Gibbs sampling, first introduced by Geman and Geman (1984), has received considerable attention in recent years, especially in Bayesian analysis, mainly due to the landmark papers of Gelfand and Smith (1990) and Gelfand et al. (1990). The method is described below.

Gibbs sampling is a Markovian updating scheme. Given an arbitrary starting set of values $U_1^{(0)}, \dots, U_m^{(0)}$, we draw $U_1^{(1)} \sim [U_1 | U_2^{(0)}, \dots, U_m^{(0)}]$, $U_2^{(1)} \sim [U_2 | U_1^{(1)}, U_3^{(0)}, \dots, U_m^{(0)}]$, $\dots$, $U_m^{(1)} \sim [U_m | U_1^{(1)}, \dots, U_{m-1}^{(1)}]$, where $[\cdot | \cdot]$ denotes the relevant conditional distributions. Hence, each variable is used and a cycle in this scheme requires generation of m random variables. After t such iterations, we get $(U_1^{(t)}, \dots, U_m^{(t)})$. As $t \rightarrow \infty$, $(U_1^{(t)}, \dots, U_m^{(t)}) \xrightarrow{d} (U_1, \dots, U_m)$. Gibbs sampling through q replications of the above t iterations generates q iid m -vectors $(U_{1j}^{(t)}, \dots, U_{mj}^{(t)})$, $j = 1, \dots, q$. $U_1, \dots, U_m$ could be real or vector valued.

Gibbs sampling consists of finding the conditional distribution of every parameter given the remaining parameters and the data. In this case the full conditionals are

given by

$$\pi(\theta_i | \theta_l (l \neq i), \alpha, \mathbf{x}) \propto \prod_{j=1}^k [F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)] g_1(\theta_i); \quad (2.3.1)$$

$$\pi(\alpha_j | \alpha_m (m \neq j), \theta, \mathbf{x}) \propto \prod_{i=1}^n [F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)] g_2(\alpha_j); \quad (2.3.2)$$

$i = 1, \dots, n$; $j = 1, \dots, k$. Note that the full conditional of θ_i does not involve the remaining $\theta_l (l \neq i)$, and the full conditional of α_j does not involve the remaining $\alpha_m (m \neq j)$.

Notice, however, that the full conditionals are also non-standard densities from which it is not possible to draw samples directly. The general procedure for generating samples in such cases is to use the Metropolis-Hastings accept-reject algorithm. If, however, F , $\bar{F}$, g_1 and g_2 are all log-concave, the full conditionals $\pi(\theta_i | \cdot)$ and $\pi(\alpha_j | \cdot)$ are all log-concave. One can then use the adaptive rejection sampling of Gilks and Wild (1992). Adaptive rejection sampling is a useful technique to sample from log-concave densities. Due to its adaptive nature it reduces the number of evaluations of the function from which the sample is drawn. This is especially useful in situations where the evaluation of the function is computationally expensive.

Adaptive rejection sampling is an improvement over general rejection sampling in the sense of computational time. Rejection sampling is a general method for sampling points independently from a density, say, $f(x)$. The density $f(x)$ need not be completely specified, i.e., rejection sampling may be performed by using $g(x)$ instead of $f(x)$, where $g(x) = cf(x)$ for some possibly unknown value of c . Rejection sampling involves calculating an envelope function of $g(x)$ which is often done by locating the supremum of $g(x)$ by using a standard optimization technique. Adaptive rejection sampling, however, avoids the need to locate the supremum of $g(x)$ by assuming log-concavity of $f(x)$. This reduces computational time considerably. It

also cuts the time down by updating the information about $g(x)$ contained in the previous step of the iteration.

To see the log-concavity of $\pi(\theta_i|\cdot)$ and $\pi(\alpha_j|\cdot)$, simply write

$$\log \pi(\theta_i|\theta_l(l \neq i), \alpha, \mathbf{x}) = \sum_{j=1}^k [x_{ij} \log F(\theta_i + \alpha_j) + (1 - x_{ij}) \log \bar{F}(\theta_i + \alpha_j)] + \log g_1(\theta_i). \quad (2.3.3)$$

If F , $\bar{F}$ and g_1 are all log-concave then clearly $\pi(\theta_i|\theta_l(l \neq i), \alpha, \mathbf{x})$ is log-concave. Similarly, if F , $\bar{F}$ and g_2 are log-concave, $\pi(\alpha_j|\alpha_m(m \neq j), \theta, \mathbf{x})$ is log-concave. The log-concavity of F and $\bar{F}$ is ensured if F is an IFR df.

It should be noted also that the full conditionals $\log \pi(\theta_i|\theta_l(l \neq i), \alpha, \mathbf{x})$ ($i = 1, \dots, n$) and $\log \pi(\alpha_j|\alpha_m(m \neq j), \theta, \mathbf{x})$ ($j = 1, \dots, k$) can all be proper, and yet the posterior $\pi(\theta, \alpha|\mathbf{x})$ can be improper (cf. Casella and George (1992)). To see this for the Rasch model, recall that the link function is $F(z) = \exp(z)/[1 + \exp(z)]$. Now, if one assigns the flat prior $\pi(\theta, \alpha) \propto 1$, then from the general result proved in Theorem 1, $\pi(\theta, \alpha|\mathbf{x})$ is improper. But

$$\begin{aligned} \pi(\theta_i|\theta_l(l \neq i), \alpha, \mathbf{x}) &\propto \prod_{j=1}^k \frac{[\exp(\theta_i + \alpha_j)x_{ij}]}{\prod_{j=1}^k [1 + \exp(\theta_i + \alpha_j)]} \\ &\propto \frac{\exp(\theta_i t_i)}{\prod_{j=1}^k [1 + \exp(\theta_i + \alpha_j)]}. \end{aligned} \quad (2.3.4)$$

Now substituting z for $\exp(\theta_i)$,

$$\begin{aligned} \int_{-\infty}^{\infty} \exp(\theta_i t_i) \prod_{j=1}^k [1 + \exp(\theta_i + \alpha_j)]^{-1} d\theta_i &= \int_0^{\infty} z^{t_i-1} \prod_{j=1}^k [1 + z \exp(\alpha_j)]^{-1} dz \\ &\leq \int_0^1 z^{t_i-1} dz + \exp(-\sum_{j=1}^k \alpha_j) \int_1^{\infty} z^{t_i-k-1} dz. \end{aligned} \quad (2.3.5)$$

Now, if $1 \leq t_i \leq k - 1$ for all $i = 1, \dots, n$,

$$\text{right hand side of (2.3.5)} = t_i^{-1} + \exp\left(-\sum_1^k \alpha_j\right)(k - t_i)^{-1} < \infty. \quad (2.3.6)$$

Similarly, if $1 \leq y_j \leq n - 1$, ($j = 1, \dots, k$),

$$\int_{-\infty}^{\infty} \pi(\alpha_j | \alpha_m (m \neq j), \theta, \mathbf{x}) d\alpha_j < \infty, \quad j = 1, \dots, k. \quad (2.3.7)$$

2.4 Analysis Based on Data For Placement Tests

The data considered in this section constitute part of the results of a mathematics placement test for entering students who satisfactorily complete two years of algebra and one year of geometry in secondary school. The data set is taken from Albert (1992). There are 200 students each answering 8 questions. Each question is classified as “correct” and “incorrect”, and is coded as 1 and 0, respectively. Our interest lies in inference about the difference of the difficulty values of these questions, or more specifically in the posterior means and posterior s.d.’s of $\alpha_j - \alpha_8$ ($j = 1, \dots, 7$). Three different links, logit, probit and log-log are considered. We consider the prior $\pi(\theta, \alpha) \propto \exp(-\sum_1^n \theta_i^2 / (2\sigma^2))$, that is θ and α are a priori independent, θ_i are *iid* $N(0, \sigma^2)$, while the α_j have flat priors. Due to the discussion of Section 2, μ can be taken as zero without loss of generality. For the Bayesian analysis, σ is taken as the tuning parameter and is used to study the sensitivity of the Bayes procedure with respect to the choice of priors.

The posterior means and s.d.’s are calculated using the Gibbs sampler (Geman and Geman, 1984, Gelfand and Smith, 1990). For this example, the number of iterates to generate a sample is taken as 50, while the number of samples is taken as 5000. A burn-in sample of 5000 is used before the actual sample is gathered. The convergence of the Gibbs sampler was checked using the Gelman-Rubin (1992) algorithm.

The usual competitors of the Bayes estimates of α are the maximum likelihood estimates for the mixed model treating θ as a random effect. As mentioned earlier in the introduction, these estimates are referred to as marginal maximum likelihood estimates (MMLE's), since they are calculated by integrating out the likelihood with respect to the assumed normal distribution of the random effect θ and maximizing the resulting "marginal likelihood" of $\mathbf{x}$ with respect to α and σ^2 . The associated standard errors for the $\alpha_j - \alpha_k$ are based on the asymptotic distribution of the MMLE's.

Alternatively, for a fixed effects formulation with the logit link, it is possible to find sufficient statistics for the nuisance parameters θ_i ($i = 1, \dots, n$). These are $T_1, \dots, T_n$, where $T_i = \sum_{j=1}^k Y_{ij}$. Hence, the conditional distribution of $\mathbf{X}$ given $\mathbf{T} = (T_1, \dots, T_n)$ does not depend on the nuisance parameter θ . The resulting likelihood, referred to as the *conditional likelihood*, depends only on α . It is shown in Andersen (1970, 1973) that the MLE of α based on this conditional likelihood (henceforth, referred to as CMLE) is consistent. Hence, for the logit link, we have considered the CMLE's as well.

The natural question is how to choose σ for the prior for the subject ability parameters. In many applications, a value would be suggested by previous work with subjects from the same or similar populations. Otherwise, in order to select a value, it can be helpful to develop a feeling for how the size of σ relates to the probabilities of a bright and not-so-bright subject getting the correct answer. The following table shows these probabilities for students at the 5th and 95th percentiles of the normal prior distribution of ability, for a variety of choices of σ and the item parameter α , for the logit link.

Table 2.1 Probabilities of Correct Response at (5th, 95th) Percentiles of Ability

	$\alpha = 0$	$\alpha = 1$	$\alpha = 2$
$\sigma = 2.0$	(.036, .964)	(.092, .986)	(.216, .995)
$\sigma = 1.5$	(.078, .922)	(.187, .870)	(.385, .989)
$\sigma = 1.0$	(.162, .838)	(.344, .934)	(.588, .975)
$\sigma = 0.5$	(.305, .695)	(.544, .861)	(.765, .944)
$\sigma = .25$	(.399, .601)	(.643, .809)	(.830, .918)
$\sigma = 0.0$	(.500, .500)	(.731, .731)	(.881, .881)

This table suggests in most applications, σ much above 1 would be quite unusual, the differences in probabilities being severe. Lacking other information, one might select $\sigma = 1.0$ for a population expected to be moderately heterogeneous and $\sigma = 0.5$ for a population expected to be mildly heterogeneous. Notice that $\sigma = 0$ corresponds to a homogeneous population.

These values apply to the logit link. The standard deviation for the cdf of the logit link equals $\pi/\sqrt{3}$ times that for probit link and $\sqrt{2}$ times that for the log-log link, which suggests corresponding standard deviation values .551 and .276 for the probit link and .707 and .354 for log-log link.

Table 2.2 displays the Bayes estimates of the item effects, with the associated standard errors in parentheses, for the logit link using $\sigma = .25, .50, 1.0$, and 1.5 . That table also displays the conditional MLEs. Tables 2.3 and 2.4 show the Bayes estimates for the probit and log-log links, using comparable values of σ for those scales. All three tables also provide the marginal MLEs, as well as the Bayes estimates for the σ values obtained via the marginal ML approach (these are 1.05, .60, and .85 for the three links). The marginal MLEs are also the posterior modes for the Bayesian approach which uses a flat prior for α . So not surprisingly, some of the marginal MLEs are very close to the Bayes estimates, especially when the corresponding σ -values are close.

In the limiting case $\sigma = 0$, observations for each item are independent Bernoulli with common success probability, and a comparison of items corresponds to a comparison of independent binomial parameters. The item effects then refer to “population-averaged” effects. As σ increases, the correlation between the responses on the various items increases. The item effects are then “subject-specific”. Subject-specific effects are typically larger than population-averaged effects, the difference increasing as the correlation increases (Neuhauser, Kalbfleisch, and Hauck, 1991). Tables 2.2-2.4 show this pattern, the Bayes estimates tending to increase as σ increases.

Table 2.2 CML, MML and Bayes estimates for logit link

i	cmle/mmle	$\hat{\alpha}_i^B - \hat{\alpha}_8^B$				
		$\sigma = 0$	$\sigma = .25$	$\sigma = .50$	$\sigma = 1.0$	$\sigma = 1.5$
1	1.991/2.036 (0.249/0.266)	1.677 (0.223)	1.709 (0.226)	1.784 (0.230)	2.012 (0.243)	2.235 (0.260)
2	3.388/3.467 (0.288/0.299)	2.892 (0.254)	2.938 (0.264)	3.063 (0.265)	3.440 (0.282)	3.787 (0.298)
3	2.869/2.934 (0.268/0.282)	2.429 (0.238)	2.466 (0.240)	2.578 (0.245)	2.911 (0.263)	3.22 (0.281)
4	2.969/3.037 (0.271/0.285)	2.518 (0.241)	2.561 (0.243)	2.671 (0.246)	3.011 (0.265)	3.325 (0.277)
5	-0.129/-0.067 (0.255/0.282)	-0.058 (0.237)	-0.055 (0.245)	-0.057 (0.244)	-0.075 (0.261)	-0.072 (0.271)
6	1.488/1.534 (0.245/0.212)	1.264 (0.217)	1.289 (0.222)	1.347 (0.224)	1.520 (0.239)	1.687 (0.257)
7	1.463/1.509 (0.244/0.264)	1.244 (0.215)	1.272 (0.223)	1.321 (0.224)	1.494 (0.240)	1.664 (0.258)

Table 2.3 MML and Bayes estimates for probit link

i	MMLE	$\hat{\alpha}_i^B - \hat{\alpha}_g^B$				
		$\sigma = 0$	$\sigma = .138$	$\sigma = .276$	$\sigma = .551$	$\sigma = .827$
1	1.170 (0.157)	1.030 (0.133)	1.049 (0.132)	1.087 (0.134)	1.201 (0.146)	1.297 (0.149)
2	1.991 (0.173)	1.743 (0.144)	1.772 (0.146)	1.834 (0.148)	2.011 (0.157)	2.167 (0.166)
3	1.699 (0.163)	1.484 (0.138)	1.506 (0.140)	1.565 (0.143)	1.727 (0.151)	1.860 (0.158)
4	1.737 (0.154)	1.535 (0.138)	1.555 (0.140)	1.608 (0.143)	1.766 (0.151)	1.901 (0.158)
5	-0.065 (0.159)	-0.033 (0.137)	-0.028 (0.138)	-0.031 (0.143)	-0.023 (0.150)	-0.020 (0.155)
6	0.875 (0.158)	0.775 (0.132)	0.789 (0.131)	0.817 (0.134)	0.907 (0.141)	0.982 (0.145)
7	0.855 (0.150)	0.761 (0.131)	0.773 (0.133)	0.805 (0.133)	0.888 (0.142)	0.960 (0.147)

Table 2.4 MML and Bayes estimates for log-log link

i	MMLE	$\hat{\alpha}_i^B - \hat{\alpha}_g^B$				
		$\sigma = 0$	$\sigma = .177$	$\sigma = .354$	$\sigma = .707$	$\sigma = 1.061$
1	1.374 (0.163)	1.114 (0.143)	1.140 (0.147)	1.202 (0.149)	1.371 (0.160)	1.518 (0.169)
2	2.508 (0.215)	2.167 (0.202)	2.202 (0.201)	2.290 (0.207)	2.508 (0.211)	2.705 (0.221)
3	2.059 (0.190)	1.747 (0.175)	1.777 (0.174)	1.854 (0.177)	2.057 (0.185)	2.236 (0.193)
4	2.180 (0.195)	1.828 (0.178)	1.865 (0.179)	1.945 (0.180)	2.179 (0.189)	2.382 (0.201)
5	0.015 (0.147)	-0.030 (0.124)	-0.023 (0.127)	-0.011 (0.132)	0.014 (0.147)	0.037 (0.157)
6	1.009 (0.155)	0.800 (0.136)	0.818 (0.138)	0.867 (0.142)	1.003 (0.151)	1.122 (0.161)
7	1.038 (0.155)	0.785 (0.135)	0.810 (0.136)	0.869 (0.138)	1.037 (0.151)	1.170 (0.159)

2.5 Example on Matched Pairs Data

In this section we consider the Bayesian analysis of another type of dependent observations, known as matched-pairs data. Dependent observations can occur in many situations. Matched-pairs design is one of them. Suppose we have n pairs of individuals, the pairing is done in such a way that the two individuals in any group tend to have similar characteristics under consideration. For instance, subjects in a pair can be matched according to their relationship to each other, such as husband-wife or father-son. One member of each pair belongs to group 1 and the other to group 2. The data from matched-pair cases are usually summarized by a 2×2 table. As introduced before in the previous section, the response variable is denoted by X_{ij} and it takes 0 for failure and 1 for success ($i = 1, \dots, n$; $j = 1, 2$). A typical table for these data looks like the following.

Table 2.5 A Typical 2×2 Table of Matched-pairs Data

	Group 2	
Group 1	1	0
1	n_{11}	n_{12}
0	n_{21}	n_{22}

where, n_{11} = number of subjects having response 1 for both the groups;

n_{12} = number of subjects having response 1 for group 1 and 0 for group 2;

n_{21} = number of subjects having response 0 for group 1 and 1 for group 2;

n_{22} = number of subjects having response 0 for both the groups;

A matched-pairs design is one of the most popular tools for case-control studies (see, for example, Liang and Zeger (1988)). We can think of group 1 as “case” and group 2 as “control” in the above table. Matched-pairs designs are widely used in fields such as ophthalmology, where an individual serves as his/her own control.

Normally for matched-pairs analysis, conditional maximum likelihood estimation is the most familiar and widely used technique. But, there are some situations where

this technique is not tenable, for example, cases where the conditioning variable (such as a sufficient statistic) is not well defined. In such cases one usually uses the marginal likelihood estimation procedure based on a mixed model. A competitor for marginal maximum likelihood estimation technique is the Bayesian procedure. In this section we will discuss three types of estimation techniques for matched-pairs cases, namely, conditional maximum likelihood, marginal maximum likelihood and Bayesian estimation. We will compare and contrast the estimates obtained from these procedures. In estimation for matched-pairs cases, the conditional and the marginal maximum likelihood estimates do not depend on the main diagonal counts, but the Bayes estimates may depend on these counts. In the next subsection, we attempt to show this dependence of Bayes estimates on the main diagonal counts.

We now illustrate the method of Sections 2.2 and 2.3 with the help of an example. The data is taken from the book 'Analysis of Binary Data' by D.R. Cox and E.J. Snell (1989). The dataset consists of 23 matched pairs of depressed patients, one member of each pair being classified as 'depersonalized' and the other as 'not depersonalized'. After treatment each patient is classified as 'recovered', coded as 1, or 'not recovered', coded as 0. The data follow.

Table 2.6 Frequency Counts of Mental Patients

Response	Response	
	Not cured (0)	Cured (1)
Not cured (0)	2	5
Cured (1)	2	14

We will be considering three link functions, namely, logit, probit and log-log for this example. First consider the following model.

$$\text{logit}(p_{ij}) = \theta_i + \alpha_j, \quad i = 1, \dots, n; \quad j = 1, 2. \quad (2.5.1)$$

Note that the above model is over-parametrized. For conditional maximum likelihood estimation we need a constraint and we use $\alpha_2 = 0$ for this purpose. For this data the Conditional Maximum Likelihood Estimate (CMLE) of α_1 and hence the difference, $\alpha_1 - \alpha_2$ is estimated to be $\log(\frac{21}{12}) = \log(\frac{7}{4}) = -0.916$ (Cox, 1958) and the asymptotic standard error is 0.837. To find the Marginal Maximum Likelihood Estimate (MMLE) we assume that the subject parameters, θ_i 's, are *iid* $N(0, \sigma^2)$. Under this setup the MMLE of interest, $\alpha_1 - \alpha_2$, is obtained using software called Mixor (Hedeker & Gibbons, 1994) and is found to be -0.916. The estimate of the prior parameter σ is 1.258. The asymptotic standard error of the MMLE is estimated to be 0.837. For binary matched pairs case there is a close connection between conditional and marginal ML estimation. In fact, Neuhaus et al. (1994) showed that under the condition for consistent estimation of the item parameters, the marginal ML estimates and the conditional ML estimates are identical. To find the Bayes estimate of the item parameters α_1 and α_2 , we consider the model described in (2.5.1) with a normal prior for the subject parameters and a flat prior for the item difficulty parameters. To avoid over-parametrization we take the mean of the normal prior to be zero (see, e.g; Albert (1992), Tsutakawa (1984)). The complete prior specification is given in the following.

$$\theta_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$\alpha_j \stackrel{iid}{\sim} \text{Uniform}(-\infty, \infty)$$

We find the Bayes estimates of α_j denoted by $\hat{\alpha}_j^B$, $j = 1, 2$ for different values of σ^2 . The result is summarized in Table 2.7. For $\sigma^2 = 1$, $\hat{\alpha}_1^B - \hat{\alpha}_2^B = -0.891$ and the absolute value of $\hat{\alpha}_1^B - \hat{\alpha}_2^B$ increases as σ^2 increases. This phenomenon is not surprising, because as σ^2 goes to ∞ the posterior of α given the data approaches the likelihood, causing $\hat{\alpha}_1^B - \hat{\alpha}_2^B$ to coincide with the ordinary ML estimate, which

is -1.833 with a standard error of 1.183. More specifically the joint posterior density of α and θ given the data $\mathbf{x}$, say $f(\alpha, \theta | \mathbf{x})$, approaches the ordinary likelihood as σ tends to ∞ . In other words,

$$\begin{aligned} f(\alpha, \theta | \mathbf{x}) &\propto \frac{e^{\sum_j y_j \alpha_j + \sum_i t_i \theta_i - \sum_i (\theta_i^2 / (2\sigma^2))}}{\prod_i \prod_j (1 + e^{\theta_i + \alpha_j})} \\ &\rightarrow \frac{e^{\sum_i t_i \theta_i + \sum_j y_j \alpha_j}}{\prod_i \prod_j (1 + e^{\theta_i + \alpha_j})}, \\ &\text{the ordinary likelihood as } \sigma \rightarrow \infty. \end{aligned}$$

It is easy to see that the result is true for any link function. This is shown in the following calculation.

$$\begin{aligned} f(\alpha, \theta | \mathbf{x}) &\propto \prod_{i=1}^n \prod_{j=1}^k F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j) e^{-\sum_i (\theta_i^2 / (2\sigma^2))} \\ &\rightarrow \prod_{i=1}^n \prod_{j=1}^k F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j), \\ &\text{the ordinary likelihood as } \sigma \rightarrow \infty. \end{aligned}$$

At this stage it is worth looking at the other extreme, i.e., σ approaching zero. Note that when $\sigma = 0$, there is no correlation between the repeated responses. In other words, we have two independent binomials for each margin. As a result of this, the model for logit link becomes,

$$\text{logit}(p_{ij}) = \alpha_j, \quad i = 1, \dots, n; \quad j = 1, 2$$

Hence, the Bayes modal estimates of α_1 and α_2 are simply,

$$\log\left(\frac{y_1}{n - y_1}\right) \text{ and } \log\left(\frac{y_2}{n - y_2}\right)$$

So,

$$\hat{\alpha}_1 - \hat{\alpha}_2 = \log\left(\frac{y_1}{n - y_1}\right) - \log\left(\frac{y_2}{n - y_2}\right)$$

$$\begin{aligned}
&= \log \left(\frac{y_1(n - y_2)}{y_2(n - y_1)} \right) \\
&= \log \left(\frac{(n_{21} + n_{22})(n_{11} + n_{21})}{(n_{12} + n_{22})(n_{11} + n_{12})} \right) \\
&= \log \left(\frac{n_{+1}n_{2+}}{n_{+2}n_{1+}} \right) \\
&= \log(\text{marginal odds ratio})
\end{aligned}$$

For $\sigma = 0$ the general model for any link function becomes,

$$F(p_{ij}) = \alpha_j, \quad i = 1, \dots, n; \quad j = 1, 2. \quad (2.5.2)$$

Under this general model, the log likelihood function is given by

$$\log L(\boldsymbol{\alpha}; \mathbf{x}) = \sum_j y_j \log F(\alpha_j) + \sum_j (n - y_j) \log \bar{F}(\alpha_j) \quad (2.5.3)$$

Hence, the Bayes modal estimate of α_j is simply,

$$\hat{\alpha}_j = F^{-1}\left(\frac{y_j}{n}\right) \quad (2.5.4)$$

Therefore, the Bayes modal estimates of $\alpha_1 - \alpha_2$ is given by

$$\hat{\alpha}_1 - \hat{\alpha}_2 = F^{-1}\left(\frac{y_1}{n}\right) - F^{-1}\left(\frac{y_2}{n}\right)$$

Now, for the logit link, under the model described in (2.5.2) the Bayes estimate (the posterior mean) of $\alpha_1 - \alpha_2$ is found to be -0.796 with a standard error of 0.743. The Bayes estimate under the original model should, therefore, go towards this value. In actual calculation we observe this pattern as we make σ smaller and smaller (Table 2.7).

Table 2.7 $\hat{\alpha}_j^B$, $\hat{\alpha}_1^B - \hat{\alpha}_2^B$ and associated Standard Errors

σ	$\hat{\alpha}_1^B$	$\hat{\alpha}_2^B$	$\hat{\alpha}_1^B - \hat{\alpha}_2^B$	$SE(\hat{\alpha}_1^B - \hat{\alpha}_2^B)$
7.07×10^{-5}	0.870	1.665	-0.794	0.746
0.224	0.893	1.700	-0.806	0.752
0.316	0.933	1.735	-0.803	0.757
0.707	1.171	2.021	-0.849	0.770
1	1.414	2.306	-0.891	0.847
1.225	1.636	2.565	-0.929	0.988
1.414	1.815	2.787	-.972	0.910
1.581	2.014	2.997	-0.983	0.935
2.236	2.681	3.791	-1.111	1.056
2.739	3.175	4.373	-1.198	1.299
3.162	3.614	4.893	-1.280	1.243
3.873	4.288	5.689	-1.401	1.356
5	2.731	4.541	-1.811	1.487

Next we consider the probit link. For this link the model becomes,

$$p_{ij} = \Phi(\theta_i + \alpha_j), \quad i = 1, \dots, n; \quad j = 1, 2.$$

First note that for this model there do not exist conditional ML estimates for the item parameters. This is because sufficient statistics for subject parameters do not exist for the above model. So for this model we will consider only marginal ML and Bayes estimates. The marginal ML estimate of $\alpha_1 - \alpha_2$ is found to be, using Mixor (Hedeker & Gibbons, 1994), $\hat{\alpha}_1 - \hat{\alpha}_2 = -0.528$ with standard error 0.476. We take, as before, θ_i 's to be *iid* $N(0, \sigma^2)$. The estimate of σ^2 from the marginal likelihood is 0.532. To find the Bayes estimates of α_1 and α_2 we take the subject parameters θ_i , $i = 1, \dots, n$ as *iid* normal with 0 mean and some variance σ^2 . As before we take a flat prior for the item difficulty parameters. The final prior specification is the following

$$\theta_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$\alpha_j \stackrel{iid}{\sim} \text{Uniform}(-\infty, \infty)$$

Let $\hat{\alpha}_{P_j}^B$ be the Bayes estimate of α_j , $j = 1, 2$. We calculate these Bayes estimates for different values of σ . The result is summarized in Table 2.8. For $\sigma = 1$, $\hat{\alpha}_{P_1}^B - \hat{\alpha}_{P_2}^B = -0.605$ and when $\sigma = 0.729$, the estimate of σ from the marginal likelihood, the Bayes estimate $\hat{\alpha}_{P_1}^B - \hat{\alpha}_{P_2}^B = -0.532$, a value which is very close to the MML estimate. From the table it is seen immediately that the estimate $\hat{\alpha}_{P_1}^B - \hat{\alpha}_{P_2}^B$ decreases as σ increases.

The unconditional maximum likelihood estimate of $\alpha_1 - \alpha_2$, for probit link, is found to be -1.132 with a standard error of 0.710. As mentioned earlier that as σ approaches infinity, the posterior goes toward the ordinary likelihood. So, the Bayes estimate of $\alpha_1 - \alpha_2$ should approach the ordinary ML estimate -1.132 as σ gets larger and larger. We see this pattern in the following table (Table 2.8). Also, when $\sigma = 0$ the model becomes the one specified in (2.5.2) with F replaced by Φ . The Bayes estimate of $\alpha_1 - \alpha_2$, under this situation, is -0.426 with a standard error of 0.421. We claim that the Bayes estimate of $\alpha_1 - \alpha_2$ under the original model should get closer to this value if we make σ smaller and smaller. We certainly observe this trend in the following table (Table 2.8).

Table 2.8 $\hat{\alpha}_{P_j}^B$, $\hat{\alpha}_{P_1}^B - \hat{\alpha}_{P_2}^B$ and associated Standard Errors

σ	$\hat{\alpha}_{P_1}^B$	$\hat{\alpha}_{P_2}^B$	$\hat{\alpha}_{P_1}^B - \hat{\alpha}_{P_2}^B$	$SE(\hat{\alpha}_{P_1}^B - \hat{\alpha}_{P_2}^B)$
7.07×10^{-4}	0.529	0.964	-0.435	0.408
1	0.717	1.322	-0.605	0.486
1.581	0.906	1.612	-0.706	0.521
2.236	1.056	1.788	-0.731	0.538
3.162	1.155	1.901	-0.745	0.531
3.873	1.163	1.906	-0.742	0.543
4.472	1.140	1.886	-0.747	0.542
5.477	1.212	1.987	-0.776	0.555
7.071	1.687	2.777	-1.090	0.667

Table 2.9 $\hat{\alpha}_{Lj}^B$, $\hat{\alpha}_{L1}^B - \hat{\alpha}_{L2}^B$ and associated Standard Errors

σ	$\hat{\alpha}_{L1}^B$	$\hat{\alpha}_{L2}^B$	$\hat{\alpha}_{L1}^B - \hat{\alpha}_{L2}^B$	$SE(\hat{\alpha}_{L1}^B - \hat{\alpha}_{L2}^B)$
0.0071	1.084	1.778	-0.694	0.658
0.071	1.086	1.791	-0.705	0.672
1	1.339	2.131	-0.792	0.695
1.581	1.566	2.515	-0.949	0.724
2.236	1.915	3.034	-1.119	0.791
2.739	2.150	3.365	-1.215	0.827
3.162	2.416	3.774	-1.358	0.878
3.873	2.696	4.197	-1.501	0.932
4.472	3.035	4.603	-1.568	0.965

Next we consider the log-log link for the same example. We follow the same steps described in the case of logit link. To find the marginal ML estimate we take θ 's to be $N(0, \sigma^2)$. The marginal ML estimate of the difference ($\hat{\alpha}_1 - \hat{\alpha}_2$) is found to be -0.748 with a standard error of 0.674. The marginal ML estimate of the prior parameter σ is found to be 1.011. To calculate Bayes estimates we take the same prior for the subject ability parameter, namely $N(0, \sigma^2)$ and a flat prior for the item parameter. More specifically,

$$\theta_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$\alpha_j \stackrel{iid}{\sim} \text{Uniform}(-\infty, \infty)$$

Let $\hat{\alpha}_{Lj}^B$ denote the Bayes estimate of α_j , $j = 1, 2$. The result is summarized in Table 2.9. For $\sigma = 1$ $\hat{\alpha}_{L1}^B - \hat{\alpha}_{L2}^B = -0.792$ and it is decreasing for increasing σ .

For log-log link, the unconditional maximum likelihood estimate of $\alpha_1 - \alpha_2$ is found to be -1.581 with a standard error of 0.881. As mentioned earlier that as σ approaches infinity, the posterior goes toward the ordinary likelihood. So, the Bayes estimate of $\alpha_1 - \alpha_2$ should get closer to the ordinary ML estimate of -1.581 for a large value of σ . We see this pattern in Table 2.9. Also, when $\sigma = 0$ the model becomes the

one specified in (2.5.2) with $F(x) = \exp(-\exp(-x))$. The Bayes estimate of $\alpha_1 - \alpha_2$, under this model, is -0.688 with a standard error of 0.656. We claim that the Bayes estimate of $\alpha_1 - \alpha_2$ under the original model should get closer to this value if we make σ smaller and smaller. We definitely observe this pattern in the above table (Table 2.9).

In this example we have seen some definite pattern in the Bayes estimates for the three link functions. For all three links, logit, probit and log-log, there is a general tendency of the Bayes estimates to decrease if we increase σ , the prior standard deviation for the subject parameters. This empirical observation on the Bayes estimates depicts a pattern which we conjectured to be true mathematically. But, we could not find a formal proof of this empirical fact as yet. In the next subsection we attempt to find a relationship of the Bayes estimates with the main diagonal counts of matched-pairs tables.

2.5.1 Effect of Diagonal Elements

The conditional ML estimates do not take into account the effect of main diagonal elements of matched pairs tables. They remain the same even if we change the main diagonal counts keeping the off diagonals the same. In this section we investigate the effect, if any, of main diagonal counts of matched-pairs tables on the Bayes estimates. For this purpose we artificially created 25 tables. The data are created in such a manner that there is an increasing trend in the counts along the main diagonal of these tables. The following table provides the datasets.

Table 2.10 Artificially created tables of Matched-pairs Data.

Table No.	Main Diagonal Counts		Off Diagonal Counts	
	n_{11}	n_{22}	n_{12}	n_{21}
1	8	20	35	10
2	8	20	36	25
3	8	20	47	48
4	8	20	49	75
5	8	20	55	195
6	15	34	35	10
7	15	34	36	25
8	15	34	47	48
9	15	34	49	75
10	15	34	55	195
11	35	85	35	10
12	35	85	36	25
13	35	85	47	48
14	35	85	49	75
15	35	85	55	195
16	50	126	35	10
17	50	126	36	25
18	50	126	47	48
19	50	126	49	75
20	50	126	55	195
21	65	135	35	10
22	65	135	36	25
23	65	135	47	48
24	65	135	49	75
25	65	135	55	195

We are interested in finding the Bayes estimates under the following model.

$$\text{logit}(p_{ij}) = \theta_i + \alpha_j, i = 1, \dots, n; \quad j = 1, 2$$

We use the same notation of Section 2.2. Our objective is to find Bayes estimates of α 's.

$$\theta_i \stackrel{iid}{\sim} N(0, \sigma^2)$$

$$\alpha_j \stackrel{iid}{\sim} \text{Uniform}(-\infty, \infty)$$

Under the above model we find that, as the main diagonal counts increase, the Bayes estimates in the first two columns also increase for each fixed off diagonal counts, whereas they are decreasing in the rest of the three columns. Our objective is to show any dependence of Bayes estimates on the main diagonal counts. This makes us expect the Bayes estimates to vary with the main diagonal counts. But at the same time we do not want these to vary so much that they are far off from the respective marginal ML estimates. This is because we know from theory that MML estimates are consistent and we want to be safe while estimating α 's employing the Bayes procedure. But Table 2.11 shows that these estimates are far off from their MML estimates (which are the same as CML estimates). We observe that as we deviate from the prior $N(0, \hat{\sigma})$, $\hat{\sigma}$ being the marginal ML estimate of the prior parameter σ , we go far away from the marginal ML estimates.

The reason behind this could be the following. The prior for the subject parameters θ_i ($i = 1, \dots, n$) is $N(0, \sigma^2)$. Now a large value of σ^2 is indicative of a huge correlation between the repeated observations. On the other hand a very low value corresponds to a weak correlation, a zero value of σ being the no correlation. So picking an arbitrary value for σ^2 could mean a wild guess on the underlying correlation structure. When we are far off from the true underlying correlation structure, Bayes estimates could end up being very different from the marginal ML estimates. In the following table, the σ -value for the estimates indicated by " $\hat{\alpha}_1 - \hat{\alpha}_2$ " is 1. We also estimated the item parameters using $\sigma = 0.224$, and the corresponding estimates are denoted by " $\hat{\alpha}_{12} - \hat{\alpha}_{22}$ ". These estimates indicate that they display similar behaviour as obtained for the case of $\sigma = 1$. We find that as the main diagonal elements increase so do the Bayes estimates for the first two columns and they tend to decrease for the rest of the three columns. We also notice that for the larger main diagonal counts the Bayes estimates usually tend to go toward the marginal ML estimates.

Table 2.11. CML and Bayes estimates for the artificially created tables.

	CML=-1.253 35 10	CML=-0.365 36 25	CML=0.021 47 48	CML=0.426 49 75	CML=1.266 55 195
8 20	$\hat{\alpha}_1-\hat{\alpha}_2$ =-1.514 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-1.732	$\hat{\alpha}_1-\hat{\alpha}_2$ =-0.514 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-0.595	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.038 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.035	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.708 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.796	$\hat{\alpha}_1-\hat{\alpha}_2$ =2.261 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =2.543
15 34	$\hat{\alpha}_1-\hat{\alpha}_2$ =-1.169 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-1.347	$\hat{\alpha}_1-\hat{\alpha}_2$ =-0.420 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-0.486	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.025 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.033	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.618 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.709	$\hat{\alpha}_1-\hat{\alpha}_2$ =2.072 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =2.348
35 85	$\hat{\alpha}_1-\hat{\alpha}_2$ =-0.370 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-0.399	$\hat{\alpha}_1-\hat{\alpha}_2$ =-0.266 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-0.308	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.023 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.020	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.457 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.526	$\hat{\alpha}_1-\hat{\alpha}_2$ =1.653 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =1.890
50 126	$\hat{\alpha}_1-\hat{\alpha}_2$ =-0.530 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-0.610	$\hat{\alpha}_1-\hat{\alpha}_2$ =-0.213 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-0.249	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.018 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.016	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.376 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.405	$\hat{\alpha}_1-\hat{\alpha}_2$ =1.440 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =1.654
65 135	$\hat{\alpha}_1-\hat{\alpha}_2$ =-0.351 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-0.530	$\hat{\alpha}_1-\hat{\alpha}_2$ =-0.184 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =-0.214	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.016 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.021	$\hat{\alpha}_1-\hat{\alpha}_2$ =0.346 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =0.398	$\hat{\alpha}_1-\hat{\alpha}_2$ =1.343 $\hat{\alpha}_{12}-\hat{\alpha}_{22}$ =1.545

It has been mentioned in the previous section that with $\sigma = 0$, the Bayes posterior mode is nothing but the marginal log odds ratio $\log[n_{+1}n_{2+}/(n_{+2}n_{1+})]$. For $\sigma > 0$, i.e., for positive correlation between the repeated responses, the above estimate is less than the subject-specific log odds ratio (Neuhaus, Kalbfleisch, and Hauck, 1991). We conjecture that if we consider tables having fixed marginal log odds ratios besides fixed off diagonal counts, we may observe some definite pattern, because this might narrow down the change in the estimates (if any) due to the change in main diagonal counts. To see this we created 8 tables, the first 4 of which have marginal log odds ratio -1.48 and the last 4 have 0.79. It was noticed that we did not have much choices for the main diagonal counts in order to keep the marginal log odds ratio, as well

as off diagonal counts fixed. One can immediately see that for the first 3 tables the total counts as well as the sum of the main diagonal counts are almost fixed. For this reason we created the fourth table where we changed all the counts keeping only the marginal log odds ratio fixed. To get more choice for the main diagonal counts we created the last four tables starting with a table having relatively large cell counts.

The results of these eight tables are summarized in Table 2.12. The value of σ was taken to be 1. We can see that the estimates are not changing considerably. This is true even for the fourth table where we have very different main diagonal counts. The same phenomenon is true for the other set of tables. As mentioned earlier that our main interest is to investigate the dependence of Bayes estimates on the main diagonal counts. So far, empirically we have noticed that the Bayes estimates are in fact changing with the main diagonal counts. But, they do not always poses the same definite pattern. We do not have concrete intuitive reason for the type of pattern we see in Table 2.11. We believe that there is some theoretical relationship of the Bayes estimates with the prior parameter σ . We also believe that the Bayes estimates and the marginal ML estimate of σ are mathematically related through the main diagonal counts. But, we have not been able to prove this formally.

Table 2.12 Artificially created tables of matched-pairs data
and associated Bayes estimates.

Table No.	Off Diagonal		Main Diagonal		$\hat{\alpha}_1^B - \hat{\alpha}_2^B$
1	35	10	8	20	-1.769
2	35	10	10	16	-1.783
3	35	10	15	11	-1.777
4	65	15	38	25	-1.776
5	55	135	195	65	0.944
6	55	135	175	65	0.972
7	55	135	170	70	0.960
8	55	135	75	160	0.955

From the preceeding discussion regarding estimation in the case of matched pairs data, it is seen that choosing a value for the prior parameter σ^2 plays an important role in the Bayesian estimation. To get around this problem of specifying an arbitrary value for σ^2 the most appropriate thing to do would be to let the data decide on the underlying correlation structure. This could be accomplished by employing hierarchical structure in the subject ability parameter. We discuss the issue of hierarchical Bayesian estimation in the next chapter.

2.6 Uniform approximation of improper priors by proper priors

So far we have considered a normal prior for the subject ability parameters and a flat prior for the item difficulty parameters. Bayesian methods depend on the prior specification. If the prior information is vague, a non-informative or a flat prior is a sensible one. Non-informative (improper) priors can be approximated by a certain class of priors. In this section we will extend the work of Mukhopadhyay and Dasgupta (1992) and Ghosh and Mukhopadhyay (1994) and obtain a class of priors to approximate improper priors for the case of item response models. We will use the notations of Section 2.2 for the parameters of the models and the likelihood function.

Consider a class of priors of the form

$$\pi_{\sigma^2}(\alpha) = \frac{c}{\sigma^{k-1}} \pi\left(\frac{\alpha}{\sigma}\right) \quad (2.6.1)$$

where c , a constant, is such that,

$$1 = \int_{\mathbb{R}^{k-1}} \frac{c}{\sigma^{k-1}} \pi\left(\frac{\alpha}{\sigma}\right) d\alpha$$

We will use the following criterion for measuring the uniform closeness of posteriors $\pi_{\sigma^2}(\alpha|\mathbf{x})$ to the posteriors $f(\alpha|\mathbf{x})$, $f(\alpha|\mathbf{x})$ being the posterior of α for a flat prior.

$$\sup_{\mathbf{x}} \int |\pi_{\sigma^2}(\alpha|\mathbf{x}) - f(\alpha|\mathbf{x})| d\alpha$$

We want to find sufficient conditions under which

$$\sup_{\mathbf{x}} \int |\pi_{\sigma^2}(\alpha|\mathbf{x}) - f(\alpha|\mathbf{x})| d\alpha \rightarrow 0 \text{ as } \sigma \rightarrow \infty. \quad (2.6.2)$$

Theorem 6.

A pair of sufficient conditions for (2.6.2) to hold is given by

$$\int_{R^{k-1}} L(\mathbf{x}, \alpha) d\alpha < \infty \quad (2.6.3)$$

$$\sup_{\mathbf{x}} \int_{R^{k-1}} \left| 1 - \pi\left(\frac{\alpha}{\sigma}\right) \right| f(\alpha|\mathbf{x}) d\mathbf{x} < g(\sigma) \rightarrow 0 \text{ as } \sigma \rightarrow \infty. \quad (2.6.4)$$

Proof.

Note that,

$$\begin{aligned} & \int |\pi_{\sigma^2}(\alpha|\mathbf{x}) - f(\alpha|\mathbf{x})| d\alpha \\ &= \int \left| \frac{L(\mathbf{x}; \alpha) \frac{e}{\sigma^{k-1}} \pi\left(\frac{\alpha}{\sigma}\right)}{\int L(\mathbf{x}; \alpha) \frac{e}{\sigma^{k-1}} \pi\left(\frac{\alpha}{\sigma}\right) d\alpha} - \frac{L(\mathbf{x}; \alpha)}{\int L(\mathbf{x}; \alpha) d\alpha} \right| d\alpha \\ &= \int \left| \frac{L(\mathbf{x}; \alpha) \pi\left(\frac{\alpha}{\sigma}\right)}{\int L(\mathbf{x}; \alpha) \pi\left(\frac{\alpha}{\sigma}\right) d\alpha} - \frac{L(\mathbf{x}; \alpha) \left(\pi\left(\frac{\alpha}{\sigma}\right) - \pi\left(\frac{\alpha}{\sigma}\right) + 1 \right)}{\int L(\mathbf{x}; \alpha) d\alpha} \right| d\alpha \\ &= \int \left| \frac{L(\mathbf{x}; \alpha) \pi\left(\frac{\alpha}{\sigma}\right)}{\int L(\mathbf{x}; \alpha) \pi\left(\frac{\alpha}{\sigma}\right) d\alpha \int L(\mathbf{x}; \alpha) d\alpha} \left(\int L(\mathbf{x}; \alpha) d\alpha - \int L(\mathbf{x}; \alpha) \pi\left(\frac{\alpha}{\sigma}\right) d\alpha \right) \right. \\ &\quad \left. - \frac{\left(1 - \pi\left(\frac{\alpha}{\sigma}\right) \right) L(\mathbf{x}; \alpha)}{\int L(\mathbf{x}; \alpha) d\alpha} \right| d\alpha \\ &\leq \int \frac{L(\mathbf{x}; \alpha) \pi\left(\frac{\alpha}{\sigma}\right)}{\int L(\mathbf{x}; \alpha) \pi\left(\frac{\alpha}{\sigma}\right) d\alpha} \left[\frac{\left| 1 - \pi\left(\frac{\alpha}{\sigma}\right) \right| L(\mathbf{x}; \alpha) d\alpha}{L(\mathbf{x}; \alpha) d\alpha} \right] d\alpha \\ &\quad + \int \frac{\left| 1 - \pi\left(\frac{\alpha}{\sigma}\right) \right| L(\mathbf{x}; \alpha)}{\int L(\mathbf{x}; \alpha) d\alpha} d\alpha \\ &= \frac{\int L(\mathbf{x}; \alpha) \pi\left(\frac{\alpha}{\sigma}\right) d\alpha}{\int L(\mathbf{x}; \alpha) \pi\left(\frac{\alpha}{\sigma}\right) d\alpha} \left[\int \left| 1 - \pi\left(\frac{\alpha}{\sigma}\right) \right| f(\alpha|\mathbf{x}) d\alpha \right] \end{aligned}$$

$$\begin{aligned}
& + \int \left| 1 - \pi\left(\frac{\alpha}{\sigma}\right) \right| f(\alpha|\mathbf{x}) d\alpha \\
& = 2 \int \left| 1 - \pi\left(\frac{\alpha}{\sigma}\right) \right| f(\alpha|\mathbf{x}) d\alpha \\
& \leq 2g(\sigma) \rightarrow 0 \text{ as } \sigma \rightarrow \infty.
\end{aligned}$$

We now check the above two conditions of the theorem for some specific link functions. First consider the logit link. Note that condition (2.6.3) has already been checked in the previous section. So we need to check the second condition. We will do that in two stages. In the first stage we will find an uniform upper bound, uniform in $\mathbf{x}$, for the numerator, and in the second stage we will find a uniform lower bound for the denominator.

First note that,

$$\begin{aligned}
& \int_{R^{k-1}} |1 - e^{\sum_j \alpha_j^2 / 2\sigma^2}| L(\mathbf{x}; \alpha) d\alpha \\
& \leq \int_{R^{k-1}} \left(\sum_j \alpha_j^2 / 2\sigma^2 \right) L(\mathbf{x}; \alpha) d\alpha \\
& = \frac{1}{2\sigma^2} \sum_{j=1}^{k-1} \int_{R^{k-1}} \alpha_j^2 \left(\int_{R^n} \frac{e^{\sum_i t_i \theta_i + \sum_j y_j \alpha_j - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i \prod_j (1 + e^{\theta_i + \alpha_j}) \prod_i (1 + e^{\theta_i})} d\theta \right) d\alpha \\
& = \frac{1}{2\sigma^2} \sum_{j=1}^{k-1} \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i (1 + e^{\theta_i})} \int_{R^{k-1}} \alpha_j^2 \frac{e^{\sum_j y_j \alpha_j}}{\prod_i \prod_j (1 + e^{\theta_i + \alpha_j})} d\alpha d\theta \\
& = \frac{1}{2\sigma^2} \sum_{l=1}^{k-1} \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i (1 + e^{\theta_i})} \left(\int_{-\infty}^{\infty} \frac{\alpha_l^2 e^{y_l \alpha_l}}{\prod_i (1 + e^{\theta_i + \alpha_l})} d\alpha_l \right) \times \\
& \quad \prod_{j(\neq l)=1}^{k-1} \left(\int_{-\infty}^{\infty} \frac{e^{y_j \alpha_j}}{\prod_i (1 + e^{\theta_i + \alpha_j})} d\alpha_j \right) d\theta \\
& \leq \frac{\Gamma(3)}{2\sigma^2} \sum_{l=1}^{k-1} \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i (1 + e^{\theta_i})} (1 + e^{-\sum_i \theta_i}) \prod_{j(\neq l)=1}^{k-1} (1 + e^{-\sum_i \theta_i}) d\theta
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{\sigma^2} \sum_{i=1}^{k-1} \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i (1 + e^{\theta_i})} (1 + e^{-\sum_i \theta_i})^{k-1} d\theta \\
&= \frac{k-1}{\sigma^2} \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i (1 + e^{\theta_i})} (1 + e^{-\sum_i \theta_i})^{k-1} d\theta \\
&\leq \frac{2^{k-2}(k-1)}{\sigma^2} \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i (1 + e^{\theta_i})} (1 + e^{-(k-1)\sum_i \theta_i}) d\theta \\
&= \frac{2^{k-2}(k-1)}{\sigma^2} \left[\prod_i \int_{-\infty}^{\infty} \frac{e^{t_i \theta - (\theta - \mu_0)^2 / 2\sigma_0^2}}{(1 + e^\theta)} d\theta + \prod_i \int_{-\infty}^{\infty} \frac{e^{(t_i - k + 1)\theta - (\theta - \mu_0)^2 / 2\sigma_0^2}}{(1 + e^\theta)} d\theta \right] \\
&\leq \frac{2^{k-2}(k-1)}{\sigma^2} \left[\prod_i \int_{-\infty}^{\infty} e^{t_i \theta - (\theta - \mu_0)^2 / 2\sigma_0^2} d\theta + \prod_i \int_{-\infty}^{\infty} e^{(t_i - k + 1)\theta - (\theta - \mu_0)^2 / 2\sigma_0^2} d\theta \right] \\
&= \frac{2^{k-2}(k-1)}{\sigma^2} \left[\prod_i \sigma_0 \sqrt{2\pi} e^{\frac{(\mu_0 + \sigma_0^2 t_i)^2 - \mu_0^2}{2\sigma_0^2}} + \prod_i \sigma_0 \sqrt{2\pi} e^{\frac{(\mu_0 + \sigma_0^2 (t_i - k + 1))^2 - \mu_0^2}{2\sigma_0^2}} \right] \\
&= \frac{2^{k-2}(k-1)}{\sigma^2} (\sigma_0 \sqrt{2\pi})^n \left[e^{n(\frac{(\mu_0 + k\sigma_0^2)^2 - \mu_0^2}{2\sigma_0^2})} + e^{n(\frac{(\mu_0 + \sigma_0^2)^2 - \mu_0^2}{2\sigma_0^2})} \right] \\
&= \frac{M}{\sigma^2} \quad (M(<\infty) \text{ independent of } \mathbf{x}) \\
&\rightarrow 0 \text{ as } \sigma \rightarrow \infty
\end{aligned}$$

Next note that,

$$\begin{aligned}
&\int_{R^{k-1}} L(\mathbf{x}; \boldsymbol{\alpha}) d\boldsymbol{\alpha} \\
&= \int_{R^{k-1}} \int_{R^n} \frac{e^{\sum_i t_i \theta_i + \sum_j y_j \alpha_j - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i \prod_j (1 + e^{\theta_i + \alpha_j}) \prod_i (1 + e^{\theta_i})} d\theta d\boldsymbol{\alpha} \\
&= \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i (1 + e^{\theta_i})} \int_{R^{k-1}} \frac{e^{\sum_j y_j \alpha_j}}{\prod_i \prod_j (1 + e^{\theta_i + \alpha_j})} d\boldsymbol{\alpha} d\theta \\
&= \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i (1 + e^{\theta_i})} \left(\prod_j \int_{-\infty}^{\infty} \frac{e^{y_j \alpha}}{\prod_i (1 + e^{\theta_i + \alpha_j})} d\alpha \right) d\theta
\end{aligned}$$

$$\begin{aligned}
&\geq \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i (1 + e^{\theta_i})} \left(\prod_j \int_{-\infty}^{\infty} \frac{e^{y_j \alpha}}{\prod_i \max(1, e^{\theta_i})(1 + e^\alpha)} d\alpha \right) d\theta \\
&= \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i \max(1, e^{\theta_i})(1 + e^{\theta_i})} \left(\prod_j \int_0^\infty \frac{u^{y_j-1}}{(1+u)^n} du \right) d\theta \\
&= \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i \max(1, e^{\theta_i})(1 + e^{\theta_i})} \left(\prod_j \int_0^1 v^{y_j-1} (1-v)^{n-y_j-1} dv \right) d\theta \\
&= \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i \max(1, e^{\theta_i})(1 + e^{\theta_i})} \left(\prod_j B(y_j, n - y_j) \right) d\theta \\
&= \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i \max(1, e^{\theta_i})(1 + e^{\theta_i})} \left(\prod_j \frac{\Gamma(y_j) \Gamma(n - y_j)}{\Gamma(n)} \right) d\theta \\
&\geq \int_{R^n} \frac{e^{\sum_i t_i \theta_i - \sum_i (\theta_i - \mu_0)^2 / 2\sigma_0^2}}{\prod_i \max(1, e^{\theta_i})(1 + e^{\theta_i})} \left(\prod_j \frac{\Gamma(1) \Gamma(1)}{\Gamma(n)} \right) d\theta \\
&= c_1 \prod_i \int_{-\infty}^{\infty} \frac{e^{t_i \theta - (\theta - \mu_0)^2 / 2\sigma_0^2}}{\max(1, e^\theta)(1 + e^\theta)} d\theta \\
&= c_1 \prod_i \left[\int_{-\infty}^0 \frac{e^{t_i \theta - (\theta - \mu_0)^2 / 2\sigma_0^2}}{\max(1, e^\theta)(1 + e^\theta)} d\theta + \int_0^\infty \frac{e^{t_i \theta - (\theta - \mu_0)^2 / 2\sigma_0^2}}{\max(1, e^\theta)(1 + e^\theta)} d\theta \right] \\
&= c_1 \prod_i \left[\int_{-\infty}^0 \frac{e^{t_i \theta - (\theta - \mu_0)^2 / 2\sigma_0^2}}{(1 + e^\theta)} d\theta + \int_0^\infty \frac{e^{t_i \theta - (\theta - \mu_0)^2 / 2\sigma_0^2}}{e^\theta (1 + e^\theta)} d\theta \right] \\
&\geq c_1 \prod_i \left[\int_{-\infty}^0 \frac{e^{k\theta - (\theta - \mu_0)^2 / 2\sigma_0^2}}{2} d\theta + \int_0^\infty \frac{e^{-(\theta - \mu_0)^2 / 2\sigma_0^2}}{2e^\theta e^\theta} d\theta \right] \\
&= (c_1/2^n) \prod_i \left[\int_{-\infty}^0 e^{k\theta - (\theta - \mu_0)^2 / 2\sigma_0^2} d\theta + \int_0^\infty e^{-(\theta - \mu_0)^2 / 2\sigma_0^2 - 2\theta} d\theta \right] \\
&= (c_1/2^n) \prod_i \left[e^{((\mu_0 + k\sigma_0^2)^2 - \mu_0^2) / 2\sigma_0^2} \int_{-\infty}^0 e^{-(\theta - (\mu_0 + k\sigma_0^2))^2 / 2\sigma_0^2} d\theta \right. \\
&\quad \left. + e^{((\mu_0 - 2\sigma_0^2)^2 - \mu_0^2) / 2\sigma_0^2} \int_0^\infty e^{-(\theta - (\mu_0 - 2\sigma_0^2))^2 / 2\sigma_0^2} d\theta \right]
\end{aligned}$$

$$\begin{aligned}
&\geq (c_1/2^n) \prod_i e^{((\mu_0-2\sigma_0^2)^2-\mu_0^2)/2\sigma_0^2} \int_0^\infty e^{-(\theta-(\mu_0-2\sigma_0^2))^2/2\sigma_0^2} d\theta \\
&= (c_1/2^n) (\sigma_0 \sqrt{2\pi})^n e^{n((\mu_0-2\sigma_0^2)^2-\mu_0^2)/2\sigma_0^2} \left[\Phi \left(\frac{\mu_0-2\sigma_0^2}{\sigma_0} \right) \right]^n
\end{aligned}$$

which is finite and independent of $\mathbf{x}$.

We next consider the log-log link. Note that,

$$\begin{aligned}
&\int_{R^{k-1}} |1 - e^{-\sum_j \alpha_j^2/2\sigma^2}| L(\mathbf{x}; \boldsymbol{\alpha}) d\boldsymbol{\alpha} \\
&\leq \int_{R^{k-1}} (\sum_j \alpha_j^2/2\sigma^2) L(\mathbf{x}; \boldsymbol{\alpha}) d\boldsymbol{\alpha} \\
&= \frac{1}{2\sigma^2} \sum_{j=1}^{k-1} \int_{R^{k-1}} \alpha_j^2 \int_{R^n} e^{-\sum_i \sum_j x_{ij} e^{-\theta_i - \alpha_j}} e^{-\sum_i x_{ik} e^{-\theta_i}} \times \\
&\quad \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \prod_i \prod_j \left(1 - e^{-e^{-\theta_i - \alpha_j}}\right)^{1-x_{ij}} d\theta d\boldsymbol{\alpha} \\
&= \frac{1}{2\sigma^2} \sum_{l=1}^{k-1} \int_{R^n} e^{-\sum_i x_{ik} e^{-\theta_i}} \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \times \\
&\quad \int_{-\infty}^\infty \alpha_l^2 e^{-\sum_i x_{il} e^{-\theta_i - \alpha_l}} \prod_i \left(1 - e^{-e^{-\theta_i - \alpha_l}}\right)^{1-x_{il}} d\alpha_l \times \\
&\quad \prod_{j(\neq l)=1}^{k-1} \int_{-\infty}^\infty e^{-\sum_i x_{ij} e^{-\theta_i - \alpha_j}} \prod_i \left(1 - e^{-e^{-\theta_i - \alpha_j}}\right)^{1-x_{ij}} d\alpha_j d\theta \\
&= \frac{1}{2\sigma^2} \sum_{l=1}^{k-1} \int_{R^n} e^{-\sum_i x_{ik} e^{-\theta_i}} \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \times \\
&\quad \left[\int_{-\infty}^0 \alpha_l^2 e^{-\sum_i x_{il} e^{-\theta_i - \alpha_l}} \prod_i \left(1 - e^{-e^{-\theta_i - \alpha_l}}\right)^{1-x_{il}} d\alpha_l \right. \\
&\quad \left. + \int_0^\infty \alpha_l^2 e^{-\sum_i x_{il} e^{-\theta_i - \alpha_l}} \prod_i \left(1 - e^{-e^{-\theta_i - \alpha_l}}\right)^{1-x_{il}} d\alpha_l \right] \times \\
&\quad \prod_{j(\neq l)=1}^{k-1} \left[\int_{-\infty}^\infty e^{-\sum_i x_{ij} e^{-\theta_i - \alpha_j}} \prod_i \left(1 - e^{-e^{-\theta_i - \alpha_j}}\right)^{1-x_{ij}} d\alpha_j d\theta \right]
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{1}{2\sigma^2} \sum_{l=1}^{k-1} \int_{R^n} e^{-\sum_i x_{ik} e^{-\theta_i}} \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \times \\
&\left[\int_{-\infty}^0 \alpha_l^2 e^{\alpha_l \sum_i x_{il} + \sum_i \theta_i x_{il}} d\alpha_l + \int_0^{\infty} \alpha_l^2 e^{-\sum_i (1-x_{il})\theta_i - \alpha_l(n-y_l)} d\alpha_l \right] \times \\
&\prod_{j(\neq l)=1}^{k-1} \left[\int_{-\infty}^{\infty} e^{-\sum_i x_{ij} e^{-\theta_i - \alpha_j}} \prod_i \left(1 - e^{-e^{-\theta_i - \alpha_j}}\right)^{1-x_{ij}} d\alpha_j d\theta \right] \\
&= \frac{1}{2\sigma^2} \sum_{l=1}^{k-1} \int_{R^n} e^{-\sum_i x_{ik} e^{-\theta_i}} \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \times \\
&\left[\frac{2e^{\sum_i \theta_i x_{il}}}{(\sum_i x_{il})^3} + \frac{2e^{-\sum_i (1-x_{il})\theta_i}}{(n-y_l)^3} \right] \times \\
&\prod_{j(\neq l)=1}^{k-1} \left[\int_{-\infty}^{\infty} e^{-\sum_i x_{ij} e^{-\theta_i - \alpha_j}} \prod_i \left(1 - e^{-e^{-\theta_i - \alpha_j}}\right)^{1-x_{ij}} d\alpha_j \right] d\theta \\
&\leq \frac{1}{\sigma^2} \sum_{l=1}^{k-1} \int_{R^n} e^{-\sum_i x_{ik} e^{-\theta_i}} \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \times \\
&\left[e^{\sum_i \theta_i x_{il}} + e^{-\sum_i (1-x_{il})\theta_i} \right] \prod_{j(\neq l)=1}^{k-1} \left[\int_{-\infty}^{\infty} e^{-\sum_i x_{ij} e^{-\theta_i - \alpha_j}} \prod_i \left(1 - e^{-e^{-\theta_i - \alpha_j}}\right)^{1-x_{ij}} d\alpha_j d\theta \right] \\
&\leq \frac{1}{\sigma^2} \sum_{l=1}^{k-1} \int_{R^n} e^{-\sum_i x_{ik} e^{-\theta_i}} \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \times \\
&\left[e^{\sum_i \theta_i x_{il}} + e^{-\sum_i (1-x_{il})\theta_i} \right] \prod_{j(\neq l)=1}^{k-1} \left[e^{\sum_i \theta_i x_{ij}} + e^{-\sum_i (1-x_{ij})\theta_i} \right] d\theta \\
&= \frac{1}{\sigma^2} \sum_{l=1}^{k-1} \int_{R^n} e^{-\sum_i x_{ik} e^{-\theta_i}} \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \times \\
&\prod_{j=1}^{k-1} \left[e^{\sum_i \theta_i x_{ij}} + e^{-\sum_i (1-x_{ij})\theta_i} \right] d\theta \\
&= \frac{k-1}{\sigma^2} \int_{R^n} e^{-\sum_i x_{ik} e^{-\theta_i}} \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} e^{\sum_i \sum_j x_{ij} \theta_i} \times
\end{aligned}$$

$$\begin{aligned}
& \prod_{j=1}^{k-1} (1 + e^{-\sum_i \theta_i}) d\theta \\
&= \frac{k-1}{\sigma^2} \int_{R^n} e^{-\sum_i x_{ik} e^{-\theta_i}} \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} \times \\
& e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} e^{\sum_i \sum_j x_{ij} \theta_i} (1 + e^{-\sum_i \theta_i})^{k-1} d\theta \\
&\leq \frac{2^{k-2}(k-1)}{\sigma^2} \left[\prod_i \int_{-\infty}^{\infty} e^{-x_{ik} e^{-\theta}} e^{-\frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} (1 - e^{-e^{-\theta}})^{1-x_{ik}} e^{t_i \theta} d\theta \right. \\
& \left. + \prod_i \int_{-\infty}^{\infty} e^{-x_{ik} e^{-\theta}} e^{-\frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} (1 - e^{-e^{-\theta}})^{1-x_{ik}} e^{(t_i - k+1)\theta} d\theta \right] \\
&\leq \frac{c_2}{\sigma^2} \left[\prod_i \int_{-\infty}^{\infty} e^{t_i \theta - \frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} d\theta + \prod_i \int_{-\infty}^{\infty} e^{(t_i - k+1)\theta - \frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} d\theta \right] \\
&= \frac{c_2}{\sigma^2} \left[\prod_i \sigma_0 \sqrt{2\pi} e^{\frac{(\mu_0 + \sigma_0^2 t_i)^2 - \mu_0^2}{2\sigma_0^2}} + \prod_i \sigma_0 \sqrt{2\pi} e^{\frac{(\mu_0 + \sigma_0^2 (t_i - k+1))^2 - \mu_0^2}{2\sigma_0^2}} \right] \\
&\leq \frac{c_2}{\sigma^2} \left[(\sigma_0 \sqrt{2\pi})^n e^{n \frac{(\mu_0 + k\sigma_0^2)^2 - \mu_0^2}{2\sigma_0^2}} + (\sigma_0 \sqrt{2\pi})^n e^{n \frac{(\mu_0 + \sigma_0^2)^2 - \mu_0^2}{2\sigma_0^2}} \right] \\
&= \frac{M_1}{\sigma^2} \rightarrow 0 \text{ as } \sigma^2 \rightarrow \infty
\end{aligned}$$

Next observe that,

$$\begin{aligned}
& \int_{R^{k-1}} L(\mathbf{x}; \boldsymbol{\alpha}) d\boldsymbol{\alpha} \\
&= \int_{R^n} \int_{R^{k-1}} e^{-\sum_i \sum_j x_{ij} e^{-\theta_i - \alpha_j}} e^{-\sum_i x_{ik} e^{-\theta_i}} \times \\
& \prod_i (1 - e^{-e^{-\theta_i}})^{1-x_{ik}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \prod_i \prod_j (1 - e^{-e^{-\theta_i - \alpha_j}})^{1-x_{ij}} d\boldsymbol{\alpha} d\theta \\
&\geq \int_{R^n} \int_{R^{k-1}} e^{-\sum_i \sum_j x_{ij} e^{-\theta_i - \alpha_j}} e^{-\sum_i x_{ik} e^{-\theta_i}} \times \\
& \prod_i (1 - e^{-e^{-\theta_i}}) e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \prod_i \prod_j (1 - e^{-e^{-\theta_i - \alpha_j}}) d\boldsymbol{\alpha} d\theta
\end{aligned}$$

$$\begin{aligned}
&\geq \int_{R^n} \int_{R^{k-1}} e^{-\sum_i \sum_j e^{-\theta_i - \alpha_j}} e^{-\sum_i e^{-\theta_i}} \times \\
&\prod_i (e^{-\theta_i} e^{-e^{-\theta_i}}) e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \prod_i \prod_j \left(e^{-\theta_i - \alpha_j} e^{-e^{-\theta_i - \alpha_j}} \right) d\alpha d\theta \\
&= \int_{R^n} e^{-\sum_i e^{-\theta_i}} e^{-\sum_i \theta_i} e^{-\sum_i e^{-\theta_i}} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \times \\
&\prod_j \int_{-\infty}^{\infty} e^{-2e^{-\alpha_j} \sum_i e^{-\theta_i}} e^{-n\alpha_j - \sum_i \theta_i} d\alpha_j d\theta \\
&= \int_{R^n} e^{-2\sum_i e^{-\theta_i}} e^{-2\sum_i \theta_i} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \prod_j \int_0^{\infty} e^{-2y \sum_i e^{-\theta_i}} y^{n-1} dy d\theta \\
&= \int_{R^n} e^{-2\sum_i e^{-\theta_i}} e^{-2\sum_i \theta_i} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \prod_j \frac{\Gamma(n)}{(2\sum_i e^{-\theta_i})^n} d\theta \\
&= \int_{R^n} e^{-2\sum_i e^{-\theta_i}} e^{-2\sum_i \theta_i} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \frac{(\Gamma(n))^{k-1}}{(2\sum_i e^{-\theta_i})^{n(k-1)}} d\theta \\
&\geq \int_{R^n} e^{-2\sum_i e^{-\theta_i}} e^{-2\sum_i \theta_i} e^{-\frac{1}{2\sigma_0^2} \sum_i (\theta_i - \mu_0)^2} \frac{1}{(e^{2\sum_i e^{-\theta_i}})^{n(k-1)}} d\theta \\
&= \prod_i \int_{-\infty}^{\infty} e^{-2e^{-\theta}} e^{-2\theta} e^{-\frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} e^{-2n(k-1)e^{-\theta}} d\theta \\
&= \prod_i \int_{-\infty}^{\infty} e^{-2\theta - \frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} e^{-(2n(k-1)+2)e^{-\theta}} d\theta \\
&= \prod_i \left[\int_{-\infty}^0 e^{-2\theta - \frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} e^{-(2n(k-1)+2)e^{-\theta}} d\theta + \int_0^{\infty} e^{-2\theta - \frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} e^{-(2n(k-1)+2)e^{-\theta}} d\theta \right] \\
&\geq \prod_i \int_0^{\infty} e^{-2\theta - \frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} e^{-(2n(k-1)+2)e^{-\theta}} d\theta \\
&\geq \prod_i \int_0^{\infty} e^{-(2n(k-1)+2)} e^{-2\theta - \frac{1}{2\sigma_0^2} (\theta - \mu_0)^2} d\theta \\
&= c_2 \left[\int_0^{\infty} e^{-\frac{1}{2\sigma_0^2} (\theta - (\mu_0 - 2\sigma_0^2))^2} e^{-\frac{(\mu_0 - 2\sigma_0^2)^2 - \mu_0^2}{2\sigma_0^2}} d\theta \right]^n
\end{aligned}$$

$$\begin{aligned}
&= c_2(\sigma_0\sqrt{2\pi})^n e^{-\frac{n((\mu_0-2\sigma_0^2)^2-\mu_0^2)}{2\sigma_0^2}} \left[\int_0^\infty \frac{1}{\sigma_0\sqrt{2\pi}} e^{-\frac{1}{2\sigma_0^2}(\theta-(\mu_0-2\sigma_0^2))^2} d\theta \right]^n \\
&= c_2(\sigma_0\sqrt{2\pi})^n e^{-\frac{n((\mu_0-2\sigma_0^2)^2-\mu_0^2)}{2\sigma_0^2}} \left[\Phi\left(\frac{\mu_0-2\sigma_0^2}{\sigma_0}\right) \right]^n
\end{aligned}$$

which is finite and independent of $\mathbf{x}$.

Lastly consider the probit link model. For this model the likelihood function is given by,

$$\begin{aligned}
L(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{x}) &\propto \prod_{i=1}^n \prod_{j=1}^{k-1} \Phi^{x_{ij}}(\theta_i + \alpha_j) \bar{\Phi}^{1-x_{ij}}(\theta_i + \alpha_j) \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \times \\
&\quad \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right)
\end{aligned} \tag{2.6.5}$$

Hence the marginal likelihood function, denoted by $L(\mathbf{x}; \boldsymbol{\alpha})$, is given by

$$L(\mathbf{x}; \boldsymbol{\alpha}) = \int_{R^n} L(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{x}) d\boldsymbol{\theta} \tag{2.6.6}$$

We note that,

$$\begin{aligned}
&\int_{R^{k-1}} |1 - e^{\sum_j \alpha_j^2 / 2\sigma^2}| L(\mathbf{x}; \boldsymbol{\alpha}) d\boldsymbol{\alpha} \\
&\leq \int_{R^{k-1}} \left(\sum_j \alpha_j^2 / 2\sigma^2 \right) L(\mathbf{x}; \boldsymbol{\alpha}) d\boldsymbol{\alpha} \\
&= \frac{1}{2\sigma^2} \sum_{j=1}^{k-1} \int_{R^{k-1}} \alpha_j^2 \left(\prod_{i=1}^n \prod_{j=1}^{k-1} \int_{R^n} \Phi^{x_{ij}}(\theta_i + \alpha_j) \bar{\Phi}^{1-x_{ij}}(\theta_i + \alpha_j) \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \right) \times \\
&\quad \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2} d\boldsymbol{\theta}\right) d\boldsymbol{\alpha} \\
&= \frac{1}{2\sigma^2} \sum_{j=1}^{k-1} \int_{R^n} \prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right) \times \\
&\quad \int_{R^{k-1}} \alpha_j^2 \prod_{i=1}^n \prod_{j=1}^{k-1} \int_{R^n} \Phi^{x_{ij}}(\theta_i + \alpha_j) \bar{\Phi}^{1-x_{ij}}(\theta_i + \alpha_j) d\boldsymbol{\alpha} d\boldsymbol{\theta}
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{2\sigma^2} \sum_{j=1}^{k-1} \int_{R^n} \prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right) \times \\
&\quad \left(\int_{-\infty}^{\infty} \alpha_j^2 \prod_{i=1}^n \Phi^{x_{ij}}(\theta_i + \alpha_j) \bar{\Phi}^{1-x_{ij}}(\theta_i + \alpha_j) d\alpha_j\right) \times \\
&\quad \prod_{l(\neq j)=1}^{k-1} \left(\int_{-\infty}^{\infty} \prod_{i=1}^n \Phi^{x_{il}}(\theta_i + \alpha_l) \bar{\Phi}^{1-x_{il}}(\theta_i + \alpha_l) d\alpha_l\right) d\theta \\
&\leq \frac{1}{2\sigma^2} \sum_{j=1}^{k-1} \int_{R^n} \prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right) \times \\
&\quad (\sqrt{2/\pi})^n \left(\int_{-\infty}^0 \alpha_j^2 e^{\sum_i x_{ij}\theta_i + \alpha_j y_j} d\alpha_j + \int_0^{\infty} \alpha_j^2 e^{-\sum_i (1-x_{ij})\theta_i - \alpha_j(n-y_j)} d\alpha_j\right) \times \\
&\quad \prod_{l(\neq j)=1}^{k-1} (\sqrt{2/\pi})^n \left(\int_{-\infty}^0 e^{\sum_i x_{il}\theta_i + \alpha_l y_l} d\alpha_l + \int_0^{\infty} e^{-\sum_i (1-x_{il})\theta_i - \alpha_l(n-y_l)} d\alpha_l\right) d\theta \\
&= \frac{1}{2\sigma^2} \sum_{j=1}^{k-1} \int_{R^n} \prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right) \times \\
&\quad (\sqrt{2/\pi})^{n(k-1)} \left(\frac{2e^{\sum_i x_{ij}\theta_i}}{y_j^3} + \frac{2e^{-\sum_i (1-x_{ij})\theta_i}}{(n-y_j)^3}\right) \times \\
&\quad \prod_{l(\neq j)=1}^{k-1} \left(\frac{e^{\sum_i x_{il}\theta_i}}{y_l} + \frac{e^{-\sum_i (1-x_{il})\theta_i}}{(n-y_l)}\right) d\theta \\
&\leq \frac{1}{2\sigma^2} \sum_{j=1}^{k-1} \int_{R^n} \prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right) \times \\
&\quad 2(\sqrt{2/\pi})^{n(k-1)} e^{\sum_i \sum_j x_{ij}\theta_i} (1 + e^{-\sum_i \theta_i})^{k-1} d\theta \\
&= \frac{2(k-1)}{2\sigma^2} (\sqrt{2/\pi})^{n(k-1)} \int_{R^n} \prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right) \times \\
&\quad e^{\sum_i \sum_{j=1}^{k-1} x_{ij}\theta_i} (1 + e^{-\sum_i \theta_i})^{k-1} d\theta \\
&\leq \frac{(k-1)}{\sigma^2} (\sqrt{2/\pi})^{n(k-1)} \int_{R^n} (\sqrt{2/\pi})^n e^{\sum_i x_{ik}\theta_i} \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right) \times
\end{aligned}$$

$$\begin{aligned}
& e^{\sum_i \sum_{j=1}^{k-1} x_{ij} \theta_i} (1 + e^{-\sum_i \theta_i})^{k-1} d\theta \\
&= \frac{(k-1)}{\sigma^2} (\sqrt{2/\pi})^{nk} \int_{R^n} e^{\sum_i t_i \theta_i - \sum_i \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}} (1 + e^{-\sum_i \theta_i})^{k-1} d\theta \\
&\leq \frac{(k-1)2^{k-1}}{\sigma^2} (\sqrt{2/\pi})^{nk} \int_{R^n} e^{\sum_i t_i \theta_i - \sum_i \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}} (1 + e^{-(k-1)\sum_i \theta_i}) d\theta \\
&= \frac{(k-1)2^{k-1}}{\sigma^2} (\sqrt{2/\pi})^{nk} \left[\prod_i \int_{-\infty}^{\infty} e^{t_i \theta - \frac{(\theta - \mu_0)^2}{2\sigma_0^2}} d\theta + \prod_i \int_{-\infty}^{\infty} e^{(t_i - k + 1)\theta - \frac{(\theta - \mu_0)^2}{2\sigma_0^2}} d\theta \right] \\
&= \frac{(k-1)2^{k-1}}{\sigma^2} (\sqrt{2/\pi})^{nk} \left[\prod_i \sigma_0 \sqrt{2\pi} e^{\frac{(\mu_0 + \sigma_0^2 t_i)^2 - \mu_0^2}{2\sigma_0^2}} + \prod_i \sigma_0 \sqrt{2\pi} e^{\frac{(\mu_0 + \sigma_0^2 (t_i - k + 1))^2 - \mu_0^2}{2\sigma_0^2}} \right] \\
&\leq \frac{(k-1)2^{k-1}}{\sigma^2} (\sqrt{2/\pi})^{nk} (\sigma_0 \sqrt{2\pi})^n \left[e^{n(\frac{(\mu_0 + k\sigma_0^2)^2 - \mu_0^2}{2\sigma_0^2})} + e^{n(\frac{(\mu_0 + \sigma_0^2)^2 - \mu_0^2}{2\sigma_0^2})} \right] \\
&= \frac{M}{\sigma^2} \quad (M(<\infty) \text{ independent of } \mathbf{x}) \\
&\rightarrow 0 \text{ as } \sigma \rightarrow \infty
\end{aligned}$$

Next we find the lower bound of $\int_{R^{k-1}} L(\alpha; \mathbf{x}) d\alpha$.

$$\begin{aligned}
& \int_{R^{k-1}} L(\alpha; \mathbf{x}) d\alpha \\
&= \int_{R^n} \int_{R^{k-1}} \prod_{i=1}^n \prod_{j=1}^{k-1} \Phi^{x_{ij}}(\theta_i + \alpha_j) \bar{\Phi}^{1-x_{ij}}(\theta_i + \alpha_j) \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \times \\
&\quad \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right) d\alpha d\theta \\
&= \int_{R^n} \left(\prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \right) \exp\left(-\sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}\right) \times \\
&\quad \prod_{j=1}^{k-1} \int_{-\infty}^{\infty} \prod_i \Phi^{x_{ij}}(\theta_i + \alpha_j) \bar{\Phi}^{1-x_{ij}}(\theta_i + \alpha_j) d\alpha_j d\theta
\end{aligned}$$

$$\begin{aligned}
&\geq \int_{R^n} \left(\prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \right) \exp \left(- \sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2} \right) \times \\
&\quad \prod_{j=1}^{k-1} \left[\int_{-\infty}^{\infty} \prod_i (1/\sqrt{2\pi}) \left(\frac{e^{2(\theta_i + \alpha_j)}}{1 + e^{4(\theta_i + \alpha_j)}} \right)^{x_{ij}} \left(\frac{e^{-2(\theta_i + \alpha_j)}}{1 + e^{-4(\theta_i + \alpha_j)}} \right)^{1-x_{ij}} d\alpha_j \right] d\theta \\
&= (1/\sqrt{2\pi})^n \int_{R^n} \left(\prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \right) \exp \left(- \sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2} \right) \times \\
&\quad \prod_{j=1}^{k-1} \left[\int_{-\infty}^{\infty} \frac{e^{2 \sum_i x_{ij}(\theta_i + \alpha_j) - 2 \sum_i (1-x_{ij})(\theta_i + \alpha_j)}}{\prod_i (1 + e^{4\theta_i + 4\alpha_j})} d\alpha_j \right] d\theta \\
&\geq \int_{R^n} (1/\sqrt{2\pi})^n \left(\prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \right) \exp \left(- \sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2} \right) \times \\
&\quad \prod_{j=1}^{k-1} \left[\int_{-\infty}^{\infty} \frac{e^{2 \sum_i x_{ij}(\theta_i + \alpha_j) - 2 \sum_i (1-x_{ij})(\theta_i + \alpha_j)}}{\prod_i \max(1, e^{4\theta_i})(1 + e^{4\alpha_j})} d\alpha_j \right] d\theta \\
&= \int_{R^n} (1/\sqrt{2\pi})^n \left(\prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \right) \exp \left(- \sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2} \right) \times \\
&\quad \frac{e^{4 \sum_i x_{ij} \theta_i}}{\prod_i \max(1, e^{4\theta_i})} \prod_{j=1}^{k-1} \left[\int_0^{\infty} \frac{u^{y_j-1}}{(1+u)^n} du \right] d\theta \\
&= \int_{R^n} (1/\sqrt{2\pi})^n \left(\prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \right) \exp \left(- \sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2} \right) \times \\
&\quad \frac{e^{4 \sum_i x_{ij} \theta_i}}{\prod_i \max(1, e^{4\theta_i})} \prod_{j=1}^{k-1} \left(\frac{\Gamma(y_j) \Gamma(n - y_j)}{\Gamma(n)} \right) d\theta \\
&\geq \int_{R^n} (1/\sqrt{2\pi})^n \left(\prod_{i=1}^n \Phi^{x_{ik}}(\theta_i) \bar{\Phi}^{1-x_{ik}}(\theta_i) \right) \exp \left(- \sum_{i=1}^n \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2} \right) \times \\
&\quad \frac{e^{4 \sum_i x_{ij} \theta_i}}{\prod_i \max(1, e^{4\theta_i})} \prod_{j=1}^{k-1} \left(\frac{\Gamma(1) \Gamma(1)}{\Gamma(n)} \right) d\theta \\
&\geq c_1 \int_{R^n} (1/\sqrt{2\pi})^n \frac{e^{2 \sum_i x_{ik} \theta_i - 2 \sum_i (1-x_{ik}) \theta_i}}{\prod_i (1 + e^{4\theta_i})} \times
\end{aligned}$$

$$\begin{aligned}
& e^{-\sum_i \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}} \frac{e^{4\sum_i \sum_{j=1}^{k-1} x_{ij} \theta_i}}{\prod_i \max(1, e^{4\theta_i})} d\theta \\
&= c_1 \int_{R^n} \frac{e^{-4\sum_i \theta_i + 4\sum_i t_i \theta_i - \sum_i \frac{(\theta_i - \mu_0)^2}{2\sigma_0^2}}}{\prod_i \max(1, e^{4\theta_i})} d\theta \\
&= c_1 \prod_{i=1}^n \left[\int_{-\infty}^0 e^{-4\theta + 4t_i \theta - \frac{(\theta - \mu_0)^2}{2\sigma_0^2}} d\theta + \int_0^{\infty} e^{-8\theta + 4t_i \theta - \frac{(\theta - \mu_0)^2}{2\sigma_0^2}} d\theta \right] \\
&\geq c_1 \prod_{i=1}^n \left[\int_{-\infty}^0 e^{4k\theta - \frac{(\theta - \mu_0)^2}{2\sigma_0^2}} d\theta + \int_0^{\infty} e^{-8\theta - \frac{(\theta - \mu_0)^2}{2\sigma_0^2}} d\theta \right]
\end{aligned}$$

which is finite and independent of $\mathbf{x}$.

Hence, for all three links we have shown that a flat prior can be uniformly approximated by any proper prior of the form given in (2.6.1).

2.7 Consistency of Marginal Maximum Likelihood Estimator

In item response literature the most widely used estimation technique is, perhaps, the conditional maximum likelihood estimation procedure. But, because of its dependence on the existence of sufficient statistics for nuisance parameters, an alternative approach is also used, which is the marginal maximum likelihood estimation procedure. The marginal maximum likelihood estimators are known to enjoy the consistency property. But, to our knowledge the formal proof of consistency for general item response models based on a parametric approach seems to be lacking. There are, however, results available on the consistency of marginal likelihood estimators based on either nonparametric or semiparametric approach. De Leeuw and Verhelst (1986) showed, based on a nonparametric approach, that for a generalized Rasch model the probability that the marginal maximum likelihood estimator and the conditional likelihood estimator are equal approaches unity as the number of subjects becomes large. Lindsay, Clogg and Grego (1991) used a semiparametric approach to arrive

at the same conclusion as De Leeuw and Verhelst (1986) but under less stringent conditions. Follman (1988) proved consistency under similar conditions as Lindsay et al. (1991) through a purely nonparametric approach. In this section we give a formal proof of consistency of marginal maximum likelihood estimators for general item response models.

We begin with the model given in (1.1.1), namely,

$$p_{ij} = F(\theta_i + \alpha_j). \quad (2.7.1)$$

Then the likelihood function is,

$$L(\theta, \alpha) = \prod_{i=1}^n \left[\left\{ \prod_{j=1}^{k-1} F^{x_{ij}}(\theta_i + \alpha_j) (1 - F(\theta_i + \alpha_j))^{1-x_{ij}} \right\} F^{x_{ik}}(\theta_i) (1 - F(\theta_i))^{1-x_{ik}} \right].$$

We assume a distribution characterized by a probability density function $\pi(\theta)$ for the subject ability parameters θ_i 's. Then the marginal likelihood for the item parameters α is given by,

$$\begin{aligned} L(\alpha) &= \prod_{i=1}^n \int \left[\prod_{j=1}^{k-1} F^{x_{ij}}(\theta + \alpha_j) (1 - F(\theta + \alpha_j))^{1-x_{ij}} \right] F^{x_{ik}}(\theta) (1 - F(\theta))^{1-x_{ik}} \pi(\theta) d\theta \\ &= \prod_{i=1}^n c_{\pi}(\mathbf{x}_i, \alpha) \quad \text{say} \end{aligned}$$

where

$$c_{\pi}(\mathbf{x}_i, \alpha) = \int \left[\prod_{j=1}^{k-1} F^{x_{ij}}(\theta + \alpha_j) (1 - F(\theta + \alpha_j))^{1-x_{ij}} \right] F^{x_{ik}}(\theta) (1 - F(\theta))^{1-x_{ik}} \pi(\theta) d\theta$$

Hence,

$$\log L(\alpha) = \sum_{i=1}^n \log c_{\pi}(\mathbf{x}_i, \alpha)$$

Let $\psi_{\pi}(\mathbf{x}, \alpha) = \left(\frac{d \log c_{\pi}(\mathbf{x}, \alpha)}{d\alpha_1}, \dots, \frac{d \log c_{\pi}(\mathbf{x}, \alpha)}{d\alpha_{k-1}} \right)'$. Then,

$$\frac{d \log L(\alpha)}{d\alpha} = \sum_{i=1}^n \psi_{\pi}(\mathbf{x}_i, \alpha)$$

Also, let

$$\begin{aligned} S_n(\alpha) &= \frac{1}{n} \frac{d \log L(\alpha)}{d\alpha} \\ &= \frac{1}{n} \sum_{i=1}^n \psi_\pi(\mathbf{x}_i, \alpha) \end{aligned}$$

The MMLE of α is the solution of the marginal likelihood equations $S_n(\alpha) = 0$. Let α_0 be the unknown true value of the parameter vector α , and Θ be an open subset of Euclidean $(k-1)$ space, $\mathbf{E}^{k-1}$.

We now state the following theorem.

Theorem 7.

Assume the following conditions.

$$(1) \quad \frac{d \log c_\pi(\mathbf{X}, \alpha)}{d\alpha_i}, \quad \frac{d^2 \log c_\pi(\mathbf{X}, \alpha)}{d\alpha_i d\alpha_j}, \quad i, j = 1, \dots, (k-1)$$

all exist and are continuous on Θ .

$$(2) \quad \text{The matrix } S'_n(\alpha_0) \text{ with the } (i, j)\text{th element given by}$$

$$\frac{1}{n} \sum_{l=1}^n \frac{d^2 \log c_\pi(\mathbf{X}_l, \alpha)}{d\alpha_i d\alpha_j} \Big|_{\alpha_0}, \quad i, j = 1, \dots, (k-1)$$

is negative definite with probability tending to unity as $n \rightarrow \infty$.

$$(3) \quad S'_n(\alpha) \text{ converges uniformly in probability to } h(\alpha) \text{ in an open neighbourhood about } \alpha_0, \text{ where the } (i, j)\text{th element of } h(\alpha) \text{ is given by,}$$

$$E \left(\frac{d^2 \log c_\pi(\mathbf{X}, \alpha)}{d\alpha_i d\alpha_j} \right) \quad i, j = 1, \dots, (k-1)$$

and

$$(4) \quad E_{\alpha_0} \left(\frac{d \log c_\pi(\mathbf{x}, \alpha)}{d\alpha} \right) \Big|_{\alpha_0} = 0$$

Then, there exists a sequence $\{\hat{\alpha}_n\}$ such that

$$S_n(\hat{\alpha}_n) = 0 \quad (2.7.2)$$

with probability tending to unity as $n \rightarrow \infty$, and

$$\hat{\alpha}_n \xrightarrow{P_{\alpha_0}} \alpha_0 \quad (2.7.3)$$

where P_{α_0} represents convergence in probability when the probability calculation is based on α_0 . If $\{\bar{\alpha}_n\}$ also satisfies (2.7.2) and (2.7.3), then $\bar{\alpha}_n = \hat{\alpha}_n$ with probability tending to unity as $n \rightarrow \infty$.

The proof of this theorem can be found in Foutz (1977).

To apply the above result in our context, we need to check the conditions under which the consistency holds. First consider the condition (4). Note that,

$$\begin{aligned} & E_{\alpha} \left(\frac{d \log c_{\pi}(X, \alpha)}{d\alpha} \middle|_{\alpha=\alpha_0} \right) \\ &= \int \left(\frac{d \log c_{\pi}(X, \alpha)}{d\alpha} \middle|_{\alpha=\alpha_0} \right) c_{\pi}(x, \alpha_0) dx \\ &= \int \frac{\left(\frac{d}{d\alpha} c_{\pi}(x, \alpha) \middle|_{\alpha=\alpha_0} \right)}{c_{\pi}(x, \alpha_0)} c_{\pi}(x, \alpha_0) dx \\ &= \int \frac{d}{d\alpha} c_{\pi}(x, \alpha) \middle|_{\alpha=\alpha_0} dx \\ &= \frac{d}{d\alpha} \left(\int c_{\pi}(x, \alpha) dx \right) \middle|_{\alpha=\alpha_0} \\ &= \frac{d}{d\alpha} (1) \middle|_{\alpha=\alpha_0} \\ &= 0 \end{aligned}$$

Next consider condition (2). Note that by Strong Law of Large Numbers (SLLN),

$$\begin{aligned} \left(\left(\frac{1}{n} \sum_{l=1}^n \frac{d^2 \log c_{\pi}(X_l, \alpha)}{d\alpha_i d\alpha_j} \right) \Big|_{\alpha=\alpha_0} \right) &\xrightarrow{wpl} \left(\left(E_{\alpha_0} \left[\frac{d^2 \log c_{\pi}(X, \alpha)}{d\alpha_i d\alpha_j} \right] \Big|_{\alpha=\alpha_0} \right) \right) \\ &= -I(\alpha_0) \\ &< 0, \end{aligned}$$

where wpl represents convergence with probability 1. Next we consider the third condition. Note that,

$$\begin{aligned} &\sup_{|\alpha-\alpha_0|<\delta} |S'_n(\alpha) - I(\alpha)| \\ &\leq \sup_{|\alpha-\alpha_0|<\delta} |S'_n(\alpha) - S'_n(\alpha_0)| + \sup_{|\alpha-\alpha_0|<\delta} |S'_n(\alpha_0) - I(\alpha_0)| \\ &\quad + \sup_{|\alpha-\alpha_0|<\delta} |I(\alpha) - I(\alpha_0)| \end{aligned} \quad (2.7.4)$$

Now the third term of (2.7.4) goes to 0 because of continuity of $I(\alpha)$. The second term of (2.7.4) goes to 0 in probability by the Weak Law of Large Numbers (WLLN). So, it is enough to show that the first term goes in probability to 0 as n approaches infinity. This result holds under certain conditions which we now state in the form of a theorem.

Theorem 8.

Suppose the following conditions hold.

$$\left. \begin{aligned} &\sup_{|\alpha-\alpha_0|<\delta} |f(\theta+\alpha) - f(\theta+\alpha_0)| < g(\theta, \delta) \rightarrow 0 \text{ as } \delta \rightarrow 0 \\ &\sup_{|\alpha-\alpha_0|<\delta} |f(\theta+\alpha)| < g_0(\theta, \delta) \\ &\sup_{|\alpha-\alpha_0|<\delta} |f^2(\theta+\alpha) - f^2(\theta+\alpha_0)| < g_1(\theta, \delta) \rightarrow 0 \text{ as } \delta \rightarrow 0 \\ &\sup_{|\alpha-\alpha_0|<\delta} |f^2(\theta+\alpha)| < g_2(\theta, \delta) \\ &\sup_{|\alpha-\alpha_0|<\delta} |f'(\theta+\alpha) - f'(\theta+\alpha_0)| < g_3(\theta, \delta) \rightarrow 0 \text{ as } \delta \rightarrow 0 \\ &\sup_{|\alpha-\alpha_0|<\delta} |f'(\theta+\alpha)| < g_4(\theta, \delta) \end{aligned} \right\} \quad (2.7.5)$$

Then the condition (3) of Theorem 7 is satisfied.

Proof.

Note that in this problem $S'_n(\alpha)$ has the following form.

$$\begin{aligned}
S'_n(\alpha) = & \frac{1}{n} \sum_{i=1}^n \frac{1}{c_\pi(\mathbf{x}, \alpha)} \int \left[\frac{X_{i1}}{F(\theta + \alpha)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha)} \right]^2 F^{X_{i1}}(\theta + \alpha) \times \\
& (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha) \pi(\theta) d\theta \\
& - \frac{1}{c_\pi(\mathbf{x}, \alpha)} \int \left[\frac{X_{i1}}{F^2(\theta + \alpha)} + \frac{1 - X_{i1}}{(1 - F(\theta + \alpha))^2} \right] F^{X_{i1}}(\theta + \alpha) \times \\
& (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f^2(\theta + \alpha) \pi(\theta) d\theta \\
& + \frac{1}{c_\pi(\mathbf{x}, \alpha)} \int \left[\frac{X_{i1}}{F(\theta + \alpha)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha)} \right] F^{X_{i1}}(\theta + \alpha) \times \\
& (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f'(\theta + \alpha) \pi(\theta) d\theta \\
& - \frac{1}{c_\pi^2(\mathbf{x}, \alpha)} \left[\int \left[\frac{X_{i1}}{F(\theta + \alpha)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha)} \right] F^{X_{i1}}(\theta + \alpha) \times \right. \\
& \left. (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha) \pi(\theta) d\theta \right]^2
\end{aligned}$$

So the expression $\sup_{|\alpha - \alpha_0| < \delta} |S'_n(\alpha) - S'_n(\alpha_0)|$ becomes,

$$\begin{aligned}
& \sup_{|\alpha - \alpha_0| < \delta} |S'_n(\alpha) - S'_n(\alpha_0)| \\
= & \frac{1}{n} \sum_{i=1}^n \left[\sup_{|\alpha - \alpha_0| < \delta} \left\{ \left| \frac{1}{c_\pi(\mathbf{x}, \alpha)} \int \left[\frac{X_{i1}}{F(\theta + \alpha)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha)} \right]^2 F^{X_{i1}}(\theta + \alpha) \times \right. \right. \right. \\
& (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha) \pi(\theta) d\theta \\
& - \frac{1}{c_\pi(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F(\theta + \alpha_0)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha_0)} \right]^2 F^{X_{i1}}(\theta + \alpha_0) \times \\
& \left. \left. (1 - F(\theta + \alpha_0))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha_0) \pi(\theta) d\theta \right\} \right]
\end{aligned}$$

$$\begin{aligned}
& + \left\{ -\frac{1}{c_\pi(\mathbf{x}, \alpha)} \int \left[\frac{X_{i1}}{F^2(\theta + \alpha)} + \frac{1 - X_{i1}}{(1 - F(\theta + \alpha))^2} \right] F^{X_{i1}}(\theta + \alpha) \times \right. \\
& (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f^2(\theta + \alpha) \pi(\theta) d\theta \\
& + \frac{1}{c_\pi(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F^2(\theta + \alpha_0)} - \frac{1 - X_{i1}}{(1 - F(\theta + \alpha_0))^2} \right] F^{X_{i1}}(\theta + \alpha_0) \times \\
& \left. (1 - F(\theta + \alpha_0))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f^2(\theta + \alpha_0) \pi(\theta) d\theta \right\} \\
& + \left\{ \frac{1}{c_\pi(\mathbf{x}, \alpha)} \int \left[\frac{X_{i1}}{F(\theta + \alpha)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha)} \right] F^{X_{i1}}(\theta + \alpha) \times \right. \\
& (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f'(\theta + \alpha) \pi(\theta) d\theta \\
& - \frac{1}{c_\pi(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F(\theta + \alpha_0)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha_0)} \right] F^{X_{i1}}(\theta + \alpha_0) \times \\
& \left. (1 - F(\theta + \alpha_0))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f'(\theta + \alpha_0) \pi(\theta) d\theta \right\} \\
& + \left\{ \frac{1}{c_\pi^2(\mathbf{x}, \alpha_0)} \left[\int \left[\frac{X_{i1}}{F(\theta + \alpha_0)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha_0)} \right] F^{X_{i1}}(\theta + \alpha_0) \times \right. \right. \\
& (1 - F(\theta + \alpha_0))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha_0) \pi(\theta) d\theta \left. \right]^2 \\
& - \frac{1}{c_\pi^2(\mathbf{x}, \alpha)} \left[\int \left[\frac{X_{i1}}{F(\theta + \alpha)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha)} \right] F^{X_{i1}}(\theta + \alpha) \times \right. \\
& \left. (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha) \pi(\theta) d\theta \right]^2 \left. \right] \quad (2.7.6)
\end{aligned}$$

Now consider the first pair of terms in (2.7.6).

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha - \alpha_0| < \delta} \left| \left\{ \frac{1}{c_\pi(\mathbf{x}, \alpha)} \int \left[\frac{X_{i1}}{F(\theta + \alpha)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha)} \right] F^{X_{i1}}(\theta + \alpha) \times \right. \right. \\
& \left. (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha) \pi(\theta) d\theta \right.
\end{aligned}$$

$$\begin{aligned}
& -\frac{1}{c_{\pi}(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F(\theta + \alpha_0)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha_0)} \right]^2 F^{X_{i1}}(\theta + \alpha_0) \times \\
& (1 - F(\theta + \alpha_0))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha_0) \pi(\theta) d\theta \Bigg\} \Bigg| \\
& \leq \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha - \alpha_0| < \delta} \left| \left(\frac{1}{c_{\pi}(\mathbf{x}, \alpha)} - \frac{1}{c_{\pi}(\mathbf{x}, \alpha_0)} \right) \int \left[\frac{X_{i1}}{F(\theta + \alpha)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha)} \right]^2 \times \right. \\
& F^{X_{i1}}(\theta + \alpha) (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha) \pi(\theta) d\theta \Bigg| \\
& + \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha - \alpha_0| < \delta} \left| \frac{1}{c_{\pi}(\mathbf{x}, \alpha_0)} \times \right. \\
& \int \left(\left[\frac{X_{i1}}{F(\theta + \alpha)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha)} \right]^2 - \left[\frac{X_{i1}}{F(\theta + \alpha_0)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha_0)} \right]^2 \right) \times \\
& F^{X_{i1}}(\theta + \alpha) (1 - F(\theta + \alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha) \pi(\theta) d\theta \Bigg| \\
& + \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha - \alpha_0| < \delta} \left| \frac{1}{c_{\pi}(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F(\theta + \alpha_0)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha_0)} \right]^2 (1 - F(\theta + \alpha))^{1-X_{i1}} \times \right. \\
& \left. \left(F^{X_{i1}}(\theta + \alpha) - F^{X_{i1}}(\theta + \alpha_0) \right) F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha) \pi(\theta) d\theta \right| \\
& + \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha - \alpha_0| < \delta} \left| \frac{1}{c_{\pi}(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F(\theta + \alpha_0)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha_0)} \right]^2 F^{X_{i1}}(\theta + \alpha_0) \times \right. \\
& \left. \left((1 - F(\theta + \alpha))^{1-X_{i1}} - (1 - F(\theta + \alpha_0))^{1-X_{i1}} \right) F^{X_{i2}}(\theta) \times \right. \\
& \left. (1 - F(\theta))^{1-X_{i2}} f(\theta + \alpha) \pi(\theta) d\theta \right| \\
& + \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha - \alpha_0| < \delta} \left| \frac{1}{c_{\pi}(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F(\theta + \alpha_0)} - \frac{1 - X_{i1}}{1 - F(\theta + \alpha_0)} \right]^2 \times \right. \\
& \left. (1 - F(\theta + \alpha_0))^{1-X_{i1}} F^{X_{i1}}(\theta + \alpha_0) (f(\theta + \alpha) - f(\theta + \alpha_0)) \times \right. \\
& \left. F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} \pi(\theta) d\theta \right|
\end{aligned} \tag{2.7.7}$$

Suppose $\sup_{|\alpha-\alpha_0|<\delta} |f(\theta+\alpha) - f(\theta+\alpha_0)|$ is bounded and goes to 0 as $\delta \rightarrow 0$. In other words if $\sup_{|\alpha-\alpha_0|<\delta} |f(\theta+\alpha) - f(\theta+\alpha_0)| \leq g(\delta)$, where $g(\delta) \rightarrow 0$ as $\delta \rightarrow 0$, then the last term of (2.7.7) converges in probability to 0, i.e.,

$$\begin{aligned} & \frac{1}{n} \sum_{i=1}^n \frac{1}{c_{\pi}(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F(\theta+\alpha_0)} - \frac{1-X_{i1}}{1-F(\theta+\alpha_0)} \right]^2 F^{X_{i1}}(\theta+\alpha_0) \times \\ & (1-F(\theta+\alpha_0))^{1-X_{i1}} \sup_{|\alpha-\alpha_0|<\delta} |f(\theta+\alpha) - f(\theta+\alpha_0)| \times \\ & F^{X_{i2}}(\theta) (1-F(\theta))^{1-X_{i2}} \pi(\theta) d\theta \\ & \xrightarrow{P_{\alpha_0}} g(\delta) E \left[\frac{1}{c_{\pi}(\mathbf{X}, \alpha_0)} \int \left[\frac{X_1}{F(\theta+\alpha_0)} - \frac{1-X_1}{1-F(\theta+\alpha_0)} \right]^2 F^{X_1}(\theta+\alpha_0) \times \right. \\ & \left. (1-F(\theta+\alpha_0))^{1-X_1} F^{X_2}(\theta) (1-F(\theta))^{1-X_2} \pi(\theta) d\theta \right] \\ & \longrightarrow 0 \quad \text{as } \delta \rightarrow 0 \end{aligned}$$

Note that the last step follows from the Lebesgue dominated convergence theorem.

Now the fourth term of (2.7.7)

$$\begin{aligned} & \leq \frac{1}{n} \sum_{i=1}^n \frac{1}{c_{\pi}(\mathbf{x}, \alpha_0)} \int \left[\frac{X_1}{F(\theta+\alpha_0)} - \frac{1-X_1}{1-F(\theta+\alpha_0)} \right]^2 \sup_{|\alpha-\alpha_0|<\delta} |f(\theta+\alpha)| \pi(\theta) \times \\ & \left((1-F(\theta+\alpha_0-\delta))^{1-X_{i1}} - (1-F(\theta+\alpha_0))^{1-X_{i1}} \right) F^{X_{i2}}(\theta) (1-F(\theta))^{1-X_{i2}} d\theta \\ & \xrightarrow{P_{\alpha_0}} E \left[\frac{1}{c_{\pi}(\mathbf{X}, \alpha_0)} \int \left[\frac{X_1}{F(\theta+\alpha_0)} - \frac{1-X_1}{1-F(\theta+\alpha_0)} \right]^2 M\pi(\theta) \times \right. \\ & \left. \left((1-F(\theta+\alpha_0-\delta))^{1-X_1} - (1-F(\theta+\alpha_0))^{1-X_1} \right) F^{X_2}(\theta) (1-F(\theta))^{1-X_2} d\theta \right] \\ & \longrightarrow 0 \quad \text{as } \delta \rightarrow 0, \end{aligned}$$

provided $\sup_{|\alpha-\alpha_0|<\delta} |f(\theta+\alpha)| < M$.

Similarly the third term of (2.7.7) converges in probability to 0 as $n \rightarrow \infty$ and $\delta \rightarrow 0$ by boundedness of $f(\theta+\alpha)$ and by continuity of $F(\cdot)$ after invoking the Lebesgue dominated convergence theorem. Proceeding exactly as above, the boundedness of $f(\theta+\alpha)$ ensures convergence of first and second terms of (2.7.7) to 0 in probability.

To show the probability convergence of the second pair of terms in (2.7.6) we break up the terms in consideration exactly as we did for the first pair of terms in (2.7.6).

This gives us the following break up.

$$\begin{aligned}
& \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha-\alpha_0|<\delta} \left| \left\{ -\frac{1}{c_\pi(\mathbf{x}, \alpha)} \int \left[\frac{X_{i1}}{F^2(\theta+\alpha)} + \frac{1-X_{i1}}{(1-F(\theta+\alpha))^2} \right] F^{X_{i1}}(\theta+\alpha) \times \right. \right. \\
& (1-F(\theta+\alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1-F(\theta))^{1-X_{i2}} f^2(\theta+\alpha) \pi(\theta) d\theta \\
& + \frac{1}{c_\pi(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F^2(\theta+\alpha_0)} + \frac{1-X_{i1}}{(1-F(\theta+\alpha_0))^2} \right] F^{X_{i1}}(\theta+\alpha_0) \times \\
& \left. \left. (1-F(\theta+\alpha_0))^{1-X_{i1}} F^{X_{i2}}(\theta) (1-F(\theta))^{1-X_{i2}} f^2(\theta+\alpha_0) \pi(\theta) d\theta \right\} \right| \\
& \leq \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha-\alpha_0|<\delta} \left| \left(\frac{1}{c_\pi(\mathbf{x}, \alpha)} - \frac{1}{c_\pi(\mathbf{x}, \alpha_0)} \right) \int \left[\frac{X_{i1}}{F^2(\theta+\alpha)} + \frac{1-X_{i1}}{(1-F(\theta+\alpha))^2} \right] \times \right. \\
& F^{X_{i1}}(\theta+\alpha) (1-F(\theta+\alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1-F(\theta))^{1-X_{i2}} f^2(\theta+\alpha) \pi(\theta) d\theta \left. \right| \\
& + \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha-\alpha_0|<\delta} \left| \frac{1}{c_\pi(\mathbf{x}, \alpha_0)} \times \right. \\
& \int \left(\left[\frac{X_{i1}}{F^2(\theta+\alpha)} + \frac{1-X_{i1}}{(1-F(\theta+\alpha))^2} \right] - \left[\frac{X_{i1}}{F^2(\theta+\alpha_0)} + \frac{1-X_{i1}}{(1-F(\theta+\alpha_0))^2} \right] \right) \times \\
& F^{X_{i1}}(\theta+\alpha) (1-F(\theta+\alpha))^{1-X_{i1}} F^{X_{i2}}(\theta) (1-F(\theta))^{1-X_{i2}} f^2(\theta+\alpha) \pi(\theta) d\theta \left. \right| \\
& + \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha-\alpha_0|<\delta} \left| \frac{1}{c_\pi(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F^2(\theta+\alpha_0)} + \frac{1-X_{i1}}{(1-F(\theta+\alpha_0))^2} \right] \times \right.
\end{aligned}$$

$$\begin{aligned}
& (1 - F(\theta + \alpha))^{1-X_{i1}} \left(F^{X_{i1}}(\theta + \alpha) - F^{X_{i1}}(\theta + \alpha_0) \right) \times \\
& \left| F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f^2(\theta + \alpha) \pi(\theta) d\theta \right| \\
& + \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha - \alpha_0| < \delta} \left| \frac{1}{c_\pi(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F^2(\theta + \alpha_0)} + \frac{1 - X_{i1}}{(1 - F(\theta + \alpha_0))^2} \right] F^{X_{i1}}(\theta + \alpha_0) \times \right. \\
& \left. \left((1 - F(\theta + \alpha))^{1-X_{i1}} - (1 - F(\theta + \alpha_0))^{1-X_{i1}} \right) \times \right. \\
& \left. F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} f^2(\theta + \alpha) \pi(\theta) d\theta \right| \\
& + \frac{1}{n} \sum_{i=1}^n \sup_{|\alpha - \alpha_0| < \delta} \left| \frac{1}{c_\pi(\mathbf{x}, \alpha_0)} \int \left[\frac{X_{i1}}{F^2(\theta + \alpha_0)} + \frac{1 - X_{i1}}{(1 - F(\theta + \alpha_0))^2} \right] \times \right. \\
& \left. (1 - F(\theta + \alpha_0))^{1-X_{i1}} F^{X_{i1}}(\theta + \alpha_0) \left(f^2(\theta + \alpha) - f^2(\theta + \alpha_0) \right) \times \right. \\
& \left. F^{X_{i2}}(\theta) (1 - F(\theta))^{1-X_{i2}} \pi(\theta) d\theta \right| \tag{2.7.8}
\end{aligned}$$

Under the assumption that $\sup_{|\alpha - \alpha_0| < \delta} |f^2(\theta + \alpha) - f^2(\theta + \alpha_0)|$ is bounded by a function not depending upon α and goes to 0 as $\delta \rightarrow 0$, the last term of (2.7.8) approaches 0 in probability as n approaches ∞ and δ goes to 0. All other terms in (2.7.8) converge in probability to 0 provided $f^2(\theta + \alpha)$ is uniformly bounded by possibly a function independent of α . The conditions needed for probability convergence of the third pair of terms in (2.7.6) are :

$$(i) \quad \sup_{|\alpha - \alpha_0| < \delta} |f'(\theta + \alpha) - f'(\theta + \alpha_0)| < g_3(\theta, \delta) \rightarrow 0 \text{ as } \delta \rightarrow 0$$

and

$$(ii) \quad \sup_{|\alpha - \alpha_0| < \delta} |f'(\theta + \alpha)| < g_4(\theta, \delta)$$

The final pair of terms in (2.7.6) can be shown to converge in probability to 0 by factoring it out in the form of $(a+b)(a-b)$ and applying the techniques and conditions used for the first pair of terms in (2.7.6). This completes the proof of the theorem.

We now check the conditions described in (2.7.5) for some commonly used link functions, viz., logit, probit and log-log links. First consider the logit link. Note that for this link function,

$$F(x) = \frac{1}{1 + e^{-x}}$$

and hence,

$$\begin{aligned} f(x) &= F(x)(1 - F(x)) \\ &\leq 1 \\ |f'(x)| &= |f(x)(1 - F(x)) - F(x)f(x)| \\ &= |F(x)(1 - F(x))(1 - 2F(x))| \\ &\leq 1 \\ f''(x) &= F(x)(1 - F(x))[1 - 6F(x) + 6F^2(x)] \\ &\leq F(x)(1 - F(x))[1 - 6F(x) + 6F(x)] \\ &\leq 1 \end{aligned}$$

Hence all the conditions listed in (2.7.5) are satisfied. So for the logit link the marginal maximum likelihood estimator of the item difficulty parameter is consistent. Next consider the probit link. For this link,

$$F(x) = \Phi(x),$$

Φ being the standard normal distribution function. Then,

$$\sup_{|\alpha - \alpha_0| < \delta} |f(\theta + \alpha) - f(\theta + \alpha_0)|$$

$$\begin{aligned}
&= \sup_{|\alpha - \alpha_0| < \delta} |e^{-(\theta + \alpha)^2/2} - e^{-(\theta + \alpha_0)^2/2}| \\
&\leq \sup_{|\alpha - \alpha_0| < \delta} |-(\theta + \alpha_0)(\alpha - \alpha_0)e^{-(\theta + \alpha_0)^2/2}| \\
&< \delta|\theta + \alpha_0| \\
&= g(\theta, \delta) \quad (\text{say})
\end{aligned} \tag{2.7.9}$$

Note that by (2.7.9) $g(\theta, \delta) \rightarrow 0$ as $\delta \rightarrow 0$. Next,

$$\begin{aligned}
&\sup_{|\alpha - \alpha_0| < \delta} |f(\theta + \alpha)| \\
&= \sup_{|\alpha - \alpha_0| < \delta} |e^{-(\theta + \alpha)^2/2}| \\
&\leq 1 \quad (\text{since } e^{-x^2} \leq 1, \forall x).
\end{aligned}$$

Also,

$$\begin{aligned}
&\sup_{|\alpha - \alpha_0| < \delta} |f^2(\theta + \alpha) - f^2(\theta + \alpha_0)| \\
&= \sup_{|\alpha - \alpha_0| < \delta} |e^{-(\theta + \alpha)^2} - e^{-(\theta + \alpha_0)^2}| \\
&\leq \sup_{|\alpha - \alpha_0| < \delta} |-2(\theta + \alpha_0)(\alpha - \alpha_0)e^{-(\theta + \alpha_0)^2}| \\
&< 2\delta|\theta + \alpha_0| \\
&\longrightarrow 0 \quad \text{as } \delta \rightarrow 0
\end{aligned} \tag{2.7.10}$$

Next, it is straightforward to see that,

$$\sup_{|\alpha - \alpha_0| < \delta} |f^2(\theta + \alpha)|$$

$$\begin{aligned}
&= \sup_{|\alpha - \alpha_0| < \delta} |e^{-(\theta + \alpha)^2}| \\
&\leq 1
\end{aligned}$$

Lastly for probit link $f'(x)$ becomes,

$$f'(x) = -xe^{-x^2/2}$$

So,

$$\begin{aligned}
&\sup_{|\alpha - \alpha_0| < \delta} |f'(\theta + \alpha) - f'(\theta + \alpha_0)| \\
&= \sup_{|\alpha - \alpha_0| < \delta} |(\theta + \alpha_0)\phi(\theta + \alpha_0) - (\theta + \alpha)\phi(\theta + \alpha)| \\
&= \sup_{|\alpha - \alpha_0| < \delta} |-(\alpha - \alpha_0)\phi(\theta + \alpha_0) - (\theta + \alpha)(\phi(\theta + \alpha) - \phi(\theta + \alpha_0))| \\
&< \delta\phi(\theta + \alpha_0) + \delta|\theta + \alpha_0|\phi(\theta + \alpha_0) \sup_{|\alpha - \alpha_0| < \delta} |\theta + \alpha| \\
&= \delta\phi(\theta + \alpha_0) + \delta|\theta + \alpha_0|\phi(\theta + \alpha_0) \sup_{|\alpha - \alpha_0| < \delta} |\theta + \alpha_0 + \alpha - \alpha_0| \\
&< \delta\phi(\theta + \alpha_0) + \delta|\theta + \alpha_0|\phi(\theta + \alpha_0)(\delta + |\theta + \alpha_0|) \\
&\longrightarrow 0 \quad \text{as } \delta \rightarrow 0
\end{aligned}$$

Also note that,

$$\begin{aligned}
&\sup_{|\alpha - \alpha_0| < \delta} |f'(\theta + \alpha)| \\
&= \sup_{|\alpha - \alpha_0| < \delta} |-(\theta + \alpha)\phi(\theta + \alpha)| \\
&\leq \sup_{|\alpha - \alpha_0| < \delta} |\theta + \alpha| \\
&= \sup_{|\alpha - \alpha_0| < \delta} |\theta + \alpha_0 + \alpha - \alpha_0| \\
&\leq \delta|\theta + \alpha_0|
\end{aligned}$$

$$\longrightarrow 0 \quad \text{as } \delta \rightarrow 0$$

We now investigate whether the condition (2.7.5) is satisfied for the log-log link.

In this case,

$$F(x) = \exp(-\exp(-x))$$

Then,

$$\begin{aligned}
 & \sup_{|\alpha - \alpha_0| < \delta} |f(\theta + \alpha) - f(\theta + \alpha_0)| \\
 = & \sup_{|\alpha - \alpha_0| < \delta} |e^{-(\theta + \alpha)} e^{-e^{-(\theta + \alpha)}} - e^{-(\theta + \alpha_0)} e^{-e^{-(\theta + \alpha_0)}}| \\
 = & \sup_{|\alpha - \alpha_0| < \delta} \left| \left(e^{-(\theta + \alpha)} - e^{-(\theta + \alpha_0)} \right) e^{-e^{-(\theta + \alpha_0)}} + e^{-(\theta + \alpha)} \left(e^{-e^{-(\theta + \alpha)}} - e^{-e^{-(\theta + \alpha_0)}} \right) \right| \\
 \leq & \sup_{|\alpha - \alpha_0| < \delta} e^{-e^{-(\theta + \alpha_0)}} \left| \left(e^{-(\theta + \alpha_0 - \delta)} - e^{-(\theta + \alpha_0)} \right) \right| \\
 & + \sup_{|\alpha - \alpha_0| < \delta} e^{-(\theta + \alpha_0 - \delta)} \left| \left(e^{-e^{-(\theta + \alpha + \delta)}} - e^{-e^{-(\theta + \alpha_0)}} \right) \right| \\
 \leq & e^{-e^{-(\theta + \alpha_0)}} e^{-(\theta + \alpha_0)} |e^\delta - 1| + e^{-(\theta + \alpha_0 - \delta)} |e^{-(\theta + \alpha_0 + \delta)} - e^{-(\theta + \alpha_0)}| \\
 \leq & e^{-e^{-(\theta + \alpha_0)}} e^{-(\theta + \alpha_0)} |e^\delta - 1| + e^{-(\theta + \alpha_0 - \delta)} e^{-(\theta + \alpha_0)} |e^\delta - 1| \quad (2.7.11) \\
 = & g(\theta, \delta) \quad (\text{say})
 \end{aligned}$$

Note that by (2.7.11) $g(\theta, \delta) \rightarrow 0$ as $\delta \rightarrow 0$. Next,

$$\begin{aligned}
 & \sup_{|\alpha - \alpha_0| < \delta} |f(\theta + \alpha)| \\
 = & \sup_{|\alpha - \alpha_0| < \delta} |e^{-(\theta + \alpha)} e^{-e^{-(\theta + \alpha)}}| \\
 \leq & e^{-(\theta + \alpha_0 - \delta)} \sup_{|\alpha - \alpha_0| < \delta} |e^{-e^{-(\theta + \alpha)}}|
 \end{aligned}$$

$$\begin{aligned}
&\leq e^{-(\theta+\alpha_0-\delta)} e^{-e^{-(\theta+\alpha_0+\delta)}} \\
&\leq 1 \quad (\text{since } xe^{-x} \leq 1, \forall x > 0).
\end{aligned}$$

Also,

$$\begin{aligned}
&\sup_{|\alpha-\alpha_0|<\delta} |f^2(\theta+\alpha) - f^2(\theta+\alpha_0)| \\
&= \sup_{|\alpha-\alpha_0|<\delta} |e^{-2(\theta+\alpha)} e^{-2e^{-(\theta+\alpha)}} - e^{-2(\theta+\alpha_0)} e^{-2e^{-(\theta+\alpha_0)}}| \\
&= \sup_{|\alpha-\alpha_0|<\delta} \left| \left(e^{-2(\theta+\alpha)} - e^{-2(\theta+\alpha_0)} \right) e^{-2e^{-(\theta+\alpha_0)}} + e^{-2(\theta+\alpha)} \left(e^{-2e^{-(\theta+\alpha)}} - e^{-2e^{-(\theta+\alpha_0)}} \right) \right| \\
&\leq e^{-2e^{-(\theta+\alpha_0)}} |e^{-2(\theta+\alpha_0-\delta)} - e^{-2(\theta+\alpha_0)}| + e^{-2(\theta+\alpha_0-\delta)} |e^{-2e^{-(\theta+\alpha_0+\delta)}} - e^{-2e^{-(\theta+\alpha_0)}}| \\
&\longrightarrow 0 \quad \text{as } \delta \rightarrow 0
\end{aligned}$$

Also, it is evident that,

$$\begin{aligned}
&\sup_{|\alpha-\alpha_0|<\delta} |f^2(\theta+\alpha)| \\
&= \sup_{|\alpha-\alpha_0|<\delta} |e^{-2(\theta+\alpha)} e^{-2e^{-(\theta+\alpha)}}| \\
&\leq \sup_{|\alpha-\alpha_0|<\delta} |e^{-2(\theta+\alpha)}| \\
&= \sup_{|\alpha-\alpha_0|<\delta} \left| \left(e^{-2(\theta+\alpha)} - e^{-2(\theta+\alpha_0)} + e^{-2(\theta+\alpha_0)} \right) \right| \\
&\leq \sup_{|\alpha-\alpha_0|<\delta} |e^{-2(\theta+\alpha)} - e^{-2(\theta+\alpha_0)}| + |e^{-2(\theta+\alpha_0)}| \\
&\leq |e^{-2(\theta+\alpha_0-\delta)} - e^{-2(\theta+\alpha_0)}| + |e^{-2(\theta+\alpha_0)}| \\
&\longrightarrow 0 \quad \text{as } \delta \rightarrow 0
\end{aligned}$$

Lastly for log-log link $f'(x)$ becomes,

$$f'(x) = e^{-x} e^{-e^{-x}} (e^{-x} - 1)$$

So,

$$\begin{aligned} & \sup_{|\alpha - \alpha_0| < \delta} |f'(\theta + \alpha) - f'(\theta + \alpha_0)| \\ \leq & \sup_{|\alpha - \alpha_0| < \delta} \left| e^{-e^{-(\theta + \alpha_0)}} \left(e^{-2(\theta + \alpha)} - e^{-2(\theta + \alpha_0)} \right) + e^{-2(\theta + \alpha)} \left(e^{-e^{-(\theta + \alpha)}} - e^{-e^{-(\theta + \alpha_0)}} \right) \right| \\ & + \sup_{|\alpha - \alpha_0| < \delta} \left| e^{-e^{-(\theta + \alpha_0)}} \left(e^{-(\theta + \alpha_0)} - e^{-(\theta + \alpha)} \right) + e^{-(\theta + \alpha)} \left(e^{-e^{-(\theta + \alpha_0)}} - e^{-e^{-(\theta + \alpha)}} \right) \right| \end{aligned} \quad (2.7.12)$$

By the previous calculation it can be shown that the expression in (2.7.12) is bounded and goes to 0 as $\delta \rightarrow 0$. Also note that,

$$\begin{aligned} & \sup_{|\alpha - \alpha_0| < \delta} |f'(\theta + \alpha)| \\ = & \sup_{|\alpha - \alpha_0| < \delta} |e^{-2(\theta + \alpha)} e^{-e^{-(\theta + \alpha)}} - e^{-(\theta + \alpha)} e^{-e^{-(\theta + \alpha)}}| \\ \leq & \sup_{|\alpha - \alpha_0| < \delta} |e^{-2(\theta + \alpha_0 - \delta)} e^{-e^{-(\theta + \alpha)}}| e^{2\delta} - e^{-2\delta}| \\ \leq & e^{-2(\theta + \alpha_0 - \delta)} e^{-e^{-(\theta + \alpha_0 + \delta)}} 4\delta e^{4\delta^2} \\ \longrightarrow & 0 \quad \text{as } \delta \rightarrow 0 \end{aligned}$$

This shows that the marginal ML estimators are consistent.

In this chapter we mainly discussed different issues of one-parameter item response models. We showed consistency of the marginal ML estimates. We obtained sufficient conditions under which improper priors can be uniformly approximated by proper priors. We developed regular Bayesian procedure to estimate item parameters in one-parameter models. We also discussed the dependence of Bayes estimates on the

standard deviation of the prior for subject parameters. Our recommendation was to use a hierarchical Bayes procedure in such cases. In the next chapter we will discuss the issues of hierarchical Bayes estimation procedure for one-parameter item response models.

CHAPTER 3

HIERARCHICAL BAYESIAN ANALYSIS OF ITEM RESPONSE MODELS FOR BINARY DATA

3.1 Introduction

In the previous chapter, we provided a subjective Bayes analysis of item response models. With a normal prior for the θ_i , this needed specification of the prior mean and the prior variance. As noted already, in the previous chapter, the posterior distribution of the contrasts remains invariant under specification of the prior mean μ . Hence, assigning a specific value to μ does not affect inference for the contrasts. However, the choice of σ^2 , the prior variance, does affect the Bayesian inference.

There are situations, however, when there is little or vague prior information about σ^2 . In such cases, assigning a specific value to σ^2 may highly bias the inference. In such situations, it may be worthwhile putting a diffuse or near diffuse prior on σ^2 , and carry out the Bayesian analysis based on such hierarchical priors.

In Section 3.2 we discuss hierarchical Bayesian models and investigate propriety of the resulting posteriors. In particular, we state and prove a theorem concerning propriety of posteriors. We also point out how intuitively appealing hierarchical models can lead to improper posteriors

Section 3.3 deals with the implementation of hierarchical Bayes methods for general one-parameter item response models. Section 3.4 contains an example for the data consisting of 200 examinees to 8 questions in a placement test in mathematics. The posterior means and the standard errors of the differences in the difficulty values

of these questions are provided for the logit, probit and log-log links. We also provide the marginal maximum likelihood estimates of these questions and compare them with the hierarchical Bayes (HB) estimates.

3.2 Bayes Procedures with Hierarchical Priors

Recall the likelihood function of (θ, α) as given in (2.2.1), namely, the following.

$$L(\theta, \alpha | \mathbf{x}) = \prod_{i=1}^n \prod_{j=1}^k \left[F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j) \right], \quad (3.2.1)$$

In the previous chapter we assumed the θ_i to be *iid* $N(\mu, \sigma^2)$, a natural class of priors for ability θ (Lord and Novick, 1968). In this situation in the first stage, suppose the θ_i are *iid* normal with mean μ and variance σ^2 , i.e.,

$$\theta_i | \mu, \sigma^2 \stackrel{iid}{\sim} N(\mu, \sigma^2) \quad i = 1, \dots, n \quad (3.2.2)$$

and α_j is flat, i.e.,

$$\alpha_j \stackrel{iid}{\sim} \text{Uniform}(-\infty, \infty) \quad j = 1, \dots, k \quad (3.2.3)$$

The conventional hyperprior for the (μ, σ^2) assumes *a priori* independence of μ and σ^2 , and

$$\mu \sim \text{Uniform}(-\infty, \infty) \quad (3.2.4)$$

and

$$r = \sigma^{-2} \sim \text{gamma}\left(\frac{1}{2}a, \frac{1}{2}b\right) \quad (3.2.5)$$

where $a > 0$ and $b + n - 1 > 0$. [A random variable V is said to follow a $\text{gamma}(a, b)$ if it has the following probability density function:

$$f(v) = \frac{a^b}{\Gamma(b)} \exp(-av) v^{b-1}.] \quad (3.2.6)$$

This allows the possibility of a negative b also. However, in this case the posterior becomes improper because of nonidentifiable likelihood as well as nonidentifiable prior.

The following theorem describes it formally.

Theorem 1.

Under the choice of priors described in (3.2.2)-(3.2.5), the posterior $f(\boldsymbol{\theta}, \boldsymbol{\alpha} | \mathbf{x})$ is improper.

Proof.

The joint posterior of $\boldsymbol{\theta}, \boldsymbol{\alpha}, \mu$ and r is

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\alpha}, \mu, r | \mathbf{x}) &\propto \prod_{i=1}^n \prod_{j=1}^k \left[F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j) \right] \times \\ &\quad r^{\frac{n}{2}} \exp \left(-\frac{r}{2} \sum_{i=1}^n (\theta_i - \mu)^2 \right) \exp \left(-\frac{1}{2} ar \right) r^{\frac{b}{2}-1} \end{aligned} \quad (3.2.7)$$

First integrating with respect to μ , one gets the joint posterior of $\boldsymbol{\theta}, \boldsymbol{\alpha}$, and r as

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\alpha}, r | \mathbf{x}) &\propto \prod_{i=1}^n \prod_{j=1}^k \left[F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j) \right] \times \\ &\quad \exp \left(-\frac{r}{2} \left(a + \sum_{i=1}^n (\theta_i - \bar{\theta})^2 \right) \right) r^{\frac{(n+b-1)}{2}-1} \end{aligned} \quad (3.2.8)$$

Now make the one-to-one transformation

$$\begin{aligned} \eta_i &= \theta_i + \alpha_k, \quad i = 1, \dots, n \\ \xi_j &= \alpha_j - \alpha_k, \quad j = 1, \dots, (k-1) \\ \alpha_k &= \alpha_k, \text{ and } r = r \end{aligned} \quad (3.2.9)$$

Writing $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{k-1})$, $\bar{\eta} = \frac{1}{n} \sum_{i=1}^n \eta_i$, the joint posterior of $\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k$ and r is given by

$$f(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k, r | \mathbf{x}) \propto \prod_{i=1}^n \prod_{j=1}^{k-1} \left[F^{x_{ij}}(\eta_i + \xi_j) \bar{F}^{1-x_{ij}}(\eta_i + \xi_j) \right] \prod_{i=1}^n F^{x_{ik}}(\eta_i) \bar{F}^{1-x_{ik}}(\eta_i) \times$$

$$\exp\left(-\frac{r}{2}\left(a + \sum_{i=1}^n (\eta_i - \bar{\eta})^2\right)\right) r^{\frac{(n+b-1)}{2}-1}, \quad (3.2.10)$$

which does not involve α_k . Hence, integrating over $\alpha_k \in (-\infty, \infty)$, one finds that the joint posterior of $(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k, r)$, and accordingly the joint posterior of $(\boldsymbol{\theta}, \boldsymbol{\alpha}, r)$, is improper. This implies the impropriety of $f(\boldsymbol{\theta}, \boldsymbol{\alpha}|\mathbf{x})$.

The impropriety of the above posterior is a consequence of the uniform distribution of the location parameter μ over $(-\infty, \infty)$. If instead one assumes μ to be known, and assigns only a distribution to r (or σ^2), the impropriety usually does not occur. Indeed, we have pointed out earlier in Chapter 2 that inference about the contrasts $\alpha_j - \alpha_m$ ($1 \leq j \neq m \leq k$) remains invariant under the choice of the location parameter μ in the general location-scale family of priors for the θ_i 's. Hence, we take $\mu = 0$, but consider the same prior for $r = \sigma^{-2}$ as given in (3.2.5). In this case the joint posterior of $\boldsymbol{\theta}, \boldsymbol{\alpha}$ and r is given by

$$\begin{aligned} f(\boldsymbol{\theta}, \boldsymbol{\alpha}, r|\mathbf{x}) &\propto \prod_{i=1}^n \prod_{j=1}^k \left[F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j) \right] \times \\ &\exp\left(-\frac{r}{2}\left(a + \sum_{i=1}^n \theta_i^2\right)\right) r^{\frac{n+b}{2}-1} \end{aligned} \quad (3.2.11)$$

With the one-to-one transformation given in (3.2.9), one gets the joint posterior of $\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k$ and r given by

$$\begin{aligned} f(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k, r|\mathbf{x}) &\propto \prod_{i=1}^n \prod_{j=1}^{k-1} \left[F^{x_{ij}}(\eta_i + \xi_j) \bar{F}^{1-x_{ij}}(\eta_i + \xi_j) \right] \prod_{i=1}^n F^{x_{ik}}(\eta_i) \bar{F}^{1-x_{ik}}(\eta_i) \times \\ &\exp\left(-\frac{r}{2}\left(a + \sum_{i=1}^n (\eta_i - \alpha_k)^2\right)\right) r^{\frac{n+b}{2}-1}. \end{aligned} \quad (3.2.12)$$

First integrating with respect to α_k , it follows from (3.2.12) that the joint posterior of $\boldsymbol{\eta}, \boldsymbol{\xi}$, and r is

$$f(\boldsymbol{\eta}, \boldsymbol{\xi}, r|\mathbf{x}) \propto \prod_{i=1}^n \prod_{j=1}^{k-1} \left[F^{x_{ij}}(\eta_i + \xi_j) \bar{F}^{1-x_{ij}}(\eta_i + \xi_j) \right] \prod_{i=1}^n F^{x_{ik}}(\eta_i) \bar{F}^{1-x_{ik}}(\eta_i) \times$$

$$\exp\left(-\frac{r}{2}\left(a + \sum_{i=1}^n (\eta_i - \bar{\eta})^2\right)\right) r^{\frac{(n+b-1)}{2}-1}. \quad (3.2.13)$$

Now integrating with respect to r , one gets the joint posterior of $(\boldsymbol{\eta}, \boldsymbol{\xi})$ given by,

$$\begin{aligned} f(\boldsymbol{\eta}, \boldsymbol{\xi} | \mathbf{x}) &\propto \prod_{i=1}^n \prod_{j=1}^{k-1} \left[F^{x_{ij}}(\eta_i + \xi_j) \bar{F}^{1-x_{ij}}(\eta_i + \xi_j) \right] \prod_{i=1}^n F^{x_{ik}}(\eta_i) \bar{F}^{1-x_{ik}}(\eta_i) \times \\ &\quad \left(a + \sum_{i=1}^n (\eta_i - \bar{\eta})^2 \right)^{-\frac{(n+b-1)}{2}} \\ &\leq a^{-\frac{b+n-1}{2}} \prod_{i=1}^n \prod_{j=1}^{k-1} \left[F^{x_{ij}}(\eta_i + \xi_j) \bar{F}^{1-x_{ij}}(\eta_i + \xi_j) \right] \times \\ &\quad \prod_{i=1}^n F^{x_{ik}}(\eta_i) \bar{F}^{1-x_{ik}}(\eta_i). \end{aligned} \quad (3.2.14)$$

Now using the arguments of Theorem 3 of the previous chapter, it follows that for the logit, probit and log-log links, the posterior of $(\boldsymbol{\eta}, \boldsymbol{\xi})$ is proper at least when $k = 2$, when $0 < y_j < n$ for all $j = 1, \dots, k$ and $t_i = 1$ for all $i = 1, \dots, n$. We conjecture that for an arbitrary k , except when $x_{i1} = \dots = x_{ik} = 0$ for at least one i or $x_{1j} = \dots = x_{nj} = 0$ for at least one j , the posterior of $(\boldsymbol{\theta}, \boldsymbol{\alpha})$ given $\mathbf{x}$ is proper under the hierarchical model where μ is fixed and σ^2 has the inverse gamma prior.

Suppose now we use a hierarchical prior for $\boldsymbol{\alpha}$. More specifically, it is assumed a priori that $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$ are independent. We assume a prior $\pi_1(\boldsymbol{\theta})$ for $\boldsymbol{\theta}$. Also, assume that conditional on (ζ, τ^2) , α_j are iid $N(\zeta, \tau^2)$, where ζ and τ^2 are marginally independent with

$$\zeta \sim \text{Uniform}(-\infty, \infty) \quad (3.2.15)$$

and

$$r = \tau^{-2} \sim \text{gamma}\left(\frac{1}{2}a, \frac{1}{2}b\right) \quad (3.2.16)$$

where $a > 0$ and $b + k - 1 > 0$. We now state and prove the following theorem which establishes the propriety of the marginal posterior of interest, i.e., $f(\alpha|\mathbf{x})$. Define

$$t_i = \sum_{j=1}^k x_{ij} \text{ and } y_j = \sum_{i=1}^n x_{ij}.$$

Theorem 2.

Under the choice of priors described in (3.2.15)-(3.2.16) for θ and α and if $1 \leq y_j \leq (n-1)$ for all $j = 1, \dots, k$, the posterior $f(\alpha|\mathbf{x})$ is proper if $\pi_1(\theta)$ has finite moment generating function when

(i) $F(x) = [1 + \exp(-x)]^{-1}$, (ii) $F(x) = \Phi(x)$, where $\Phi(\cdot)$ is the standard normal distribution function, and (iii) $F(x) = \exp(-\exp(-x))$.

Proof.

From (3.2.1) the likelihood function for any unspecified distribution function $F(\cdot)$ is given by,

$$L(\theta, \alpha|\mathbf{x}) = \prod_{i=1}^n \prod_{j=1}^k \left[F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j) \right], \quad (3.2.17)$$

Under the given prior the joint posterior of θ , α , ζ and r is given by,

$$f(\theta, \alpha, \zeta, r|\mathbf{x}) \propto L(\theta, \alpha|\mathbf{x}) \exp\left(-\frac{r}{2} \sum_{j=1}^k (\alpha_j - \zeta)^2\right) \exp\left(-\frac{1}{2}ar\right) r^{\frac{1}{2}(b+k)-1} \pi_1(\theta) \quad (3.2.18)$$

Integrating (3.2.18) with respect to ζ we get,

$$f(\theta, \alpha, r|\mathbf{x}) \propto L(\theta, \alpha|\mathbf{x}) \exp\left(-\frac{r}{2} \sum_{j=1}^k (\alpha_j - \bar{\alpha})^2\right) \exp\left(-\frac{1}{2}ar\right) r^{\frac{1}{2}(b+k)-1} \pi_1(\theta) \quad (3.2.19)$$

Finally, integrating (3.2.19) with respect to r we get,

$$\begin{aligned} f(\theta, \alpha|\mathbf{x}) &\propto L(\theta, \alpha|\mathbf{x}) \times \left[a + \sum_{j=1}^k (\alpha_j - \bar{\alpha})^2 \right]^{-\frac{(b+k-1)}{2}} \pi_1(\theta) \\ &\leq L(\theta, \alpha|\mathbf{x}) a^{-\frac{(b+k-1)}{2}} \pi_1(\theta) \end{aligned} \quad (3.2.20)$$

Now following the steps of the proof of Theorem 4 of the previous chapter, it can be shown that the posterior is proper whenever $\pi_1(\theta)$ has finite moment generating function.

3.3 Implementation Of Bayes Procedures

Consider the general link $p_{ij} = F(\theta_i + \alpha_j)$, $i = 1, \dots, n$, $j = 1, \dots, k$, and the prior $g(\theta, \alpha) \propto \prod_{i=1}^n \pi_1(\theta_i | \mu, \sigma^2) \pi_3(\sigma^2) \prod_{j=1}^k \pi_2(\alpha_j)$. Both π_2 and π_3 can be improper as long as the posterior $f(\theta, \alpha | \mathbf{x})$ remains proper. As indicated earlier in the previous chapter, we shall use Gibbs sampling, introduced by Geman and Geman (1984), to generate samples from the marginal densities.

Gibbs sampling consists of finding the conditional distribution of every parameter given the remaining parameters and the data. In this case the full conditionals are given by

$$f(\theta_i | \theta_l (l \neq i), \sigma^2, \alpha, \mathbf{x}) \propto \prod_{j=1}^k [F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)] \pi_1(\theta_i | \mu, \sigma^2); \quad (3.3.1)$$

$$f(\alpha_j | \alpha_m (m \neq j), \sigma^2, \theta, \mathbf{x}) \propto \prod_{i=1}^n [F^{x_{ij}}(\theta_i + \alpha_j) \bar{F}^{1-x_{ij}}(\theta_i + \alpha_j)] \pi_2(\alpha_j); \quad (3.3.2)$$

$$f(\sigma^2 | \theta, \alpha, \mathbf{x}) \propto \prod_{i=1}^n [\pi_1(\theta_i | \mu, \sigma^2) \pi_3(\sigma^2)]; \quad (3.3.3)$$

$i = 1, \dots, n$; $j = 1, \dots, k$. Note that the full conditional of θ_i does not involve the remaining $\theta_l (l \neq i)$, and the full conditional of α_j does not involve the remaining $\alpha_m (m \neq j)$.

We will be using the Adaptive Rejection sampling of Gilks and Wild (1992) to generate samples from the marginal densities provided the full conditionals are all log-concave densities. To establish this log-concavity of the full conditionals before we use the Adaptive Rejection sampling, we state and prove the following theorem.

Theorem 3.

If F , $\bar{F}$, π_1 , π_2 and π_3 are log-concave densities, then the full conditionals, $f(\theta_i|\cdot)$, $f(\alpha_j|\cdot)$ and $f(\sigma^2|\cdot)$ are all log-concave.

Proof.

To see the log-concavity of $f(\theta_i|\cdot)$, $f(\alpha_j|\cdot)$ and $f(\sigma^2|\cdot)$ let us write the following.

$$\log f(\theta_i|\theta_l(l \neq i), \sigma^2, \boldsymbol{\alpha}, \mathbf{x}) = \sum_{j=1}^k [x_{ij} \log F(\theta_i + \alpha_j) + (1 - x_{ij}) \log \bar{F}(\theta_i + \alpha_j)] + \log \pi_1(\theta_i|\mu, \sigma^2) \quad (3.3.4)$$

If F , $\bar{F}$ and π_1 are all log-concave then clearly $f(\theta_i|\theta_l(l \neq i), \sigma^2, \boldsymbol{\alpha}, \mathbf{x})$ is log-concave.

Similarly,

$$\log f(\alpha_j|\boldsymbol{\theta}, \alpha_m(m \neq j), \sigma^2, \mathbf{x}) = \sum_{i=1}^n [x_{ij} \log F(\theta_i + \alpha_j) + (1 - x_{ij}) \log \bar{F}(\theta_i + \alpha_j)] + \log \pi_2(\alpha_j) \quad (3.3.5)$$

Hence, if F , $\bar{F}$ and π_2 are log-concave, $f(\alpha_j|\alpha_m(m \neq j), \sigma^2, \boldsymbol{\theta}, \mathbf{x})$ is log-concave. The log-concavity of $f(\sigma^2|\boldsymbol{\theta}, \boldsymbol{\alpha}, \mathbf{x})$ follows similarly.

3.4 An Example

The data considered in this section is the same mathematics placement test dataset that we have considered in the previous chapter. Altogether there are 32 questions, but we consider only the first 8. Our interest lies in inference about the difference of the difficulty values of these questions, or more specifically in the posterior means and posterior s.d.'s of $\alpha_j - \alpha_8$ ($j = 1, \dots, 7$). Three different links, logit, probit and log-log are considered. We consider the following hierarchical prior.

$$\theta_i|\sigma^2 \stackrel{iid}{\sim} N(0, \sigma^2) \quad i = 1, \dots, n$$

$$\alpha_j \stackrel{iid}{\sim} \text{Uniform}(-\infty, \infty) \quad j = 1, \dots, k$$

$$r = \sigma^{-2} \sim \text{gamma}(\frac{1}{2}a, \frac{1}{2}b) \quad (3.4.1)$$

Here a^{-1} represents the scale parameter and b represents the shape parameter of the distribution of r . The mean and variance of this distribution are b/a and $2b/a^2$ respectively. Therefore, it can be seen that

$$b = \frac{2}{(\text{cv})^2}$$

where cv denotes the coefficient of variation of the distribution of r . Hence, a/b represents a guess at the location of σ^2 and b measures the prior precision of this guess. So, to model the uncertainty about σ^2 one should assign a very small value of b including 0. A value of 0 for b implies a noninformative prior for σ^2 . Since, using $b = 0$ does not pose problem in getting a proper posterior we decided to take this value of b . We discuss the choice of a at the end of this section.

The estimate of σ for the logit link from the marginal likelihood is found to be 0.891 with standard error 0.109 and the HB estimate of σ is 0.881 with standard error 0.090. For the probit link the estimate of σ from the marginal likelihood is given by 0.511 with standard error 0.049, whereas the HB estimates of σ for this link is given by 0.510 with standard error 0.043. The estimate of σ for the log-log link obtained from the marginal likelihood is given by 0.606 with standard error 0.072 and the corresponding HB estimate is 0.589 with standard error 0.064.

We have noticed in the previous chapter that, when σ is close to the value of its marginal ML estimate, the Bayes estimates are close to the corresponding marginal ML estimates. However, it is reasonable to believe that if we let the data decide about a suitable value for σ , rather than putting some arbitrary number, we will get better estimates. This naturally leads to a hierarchical model where we put a distribution on σ^2 . By doing so, we get the hierarchical Bayes estimates of the item parameters. These estimates are provided in the Tables 3.1-3.3. We can see from these tables

that the HB estimates of the item difficulty parameters are reasonably close to their marginal ML estimates. This is because the HB estimate of σ is very close to its marginal ML estimate, and this phenomenon is true for all three links. Also, we checked the robustness of the HB estimates with respect to the choice of a for the Gamma distribution. The natural question is how to select a . The usual practice is to use the value 0 for a giving a noninformative prior for σ which reflects the uncertainty in the prior selection. But, as we mentioned earlier, in our case we could not do that for technical reason, because otherwise the posterior becomes improper. As the tables show, the HB estimates do not change much (in fact, very little) if we vary a from 1 to 0.001. So, our recommendation is to use a positive number close to zero.

Table 3.1 MML and HB estimates for logit link

i	MMLE	$\hat{\alpha}_i^{HB} - \hat{\alpha}_8^{HB}$			
		$a = 1$	$a = .5$	$a = 0.01$	$a = 0.001$
1	0.739 (0.179)	0.753 (0.182)	0.743 (0.179)	0.745 (0.181)	0.749 (0.182)
2	2.163 (0.236)	2.199 (0.238)	2.178 (0.238)	2.189 (0.241)	2.182 (0.239)
3	1.626 (0.207)	1.651 (0.211)	1.636 (0.210)	1.641 (0.207)	1.640 (0.211)
4	1.729 (0.212)	1.754 (0.215)	1.738 (0.212)	1.751 (0.214)	1.744 (0.214)
5	-1.376 (0.200)	-1.386 (0.202)	-1.386 (0.201)	-1.383 (0.200)	-1.379 (0.204)
6	0.245 (0.172)	0.252 (0.174)	0.245 (0.175)	0.253 (0.176)	0.252 (0.172)
7	0.220 (0.172)	0.229 (0.173)	0.225 (0.172)	0.228 (0.173)	0.227 (0.175)

Table 3.2 MML and HB estimates for probit link

i	MMLE	$\hat{\alpha}_i^{HB} - \hat{\alpha}_8^{HB}$			
		$a = 1$	$a = .5$	$a = 0.01$	$a = 0.001$
1	0.434 (0.102)	0.453 (0.106)	0.451 (0.108)	0.438 (0.106)	0.434 (0.108)
2	1.259 (0.122)	1.284 (0.130)	1.283 (0.129)	1.270 (0.130)	1.267 (0.131)
3	0.963 (0.112)	0.987 (0.122)	0.981 (0.120)	0.972 (0.120)	0.968 (0.120)
4	1.004 (0.112)	1.024 (0.117)	1.028 (0.120)	1.014 (0.117)	1.007 (0.119)
5	-0.819 (0.114)	-0.817 (0.115)	-0.818 (0.117)	-0.822 (0.116)	-0.822 (0.114)
6	0.138 (0.101)	0.151 (0.105)	0.149 (0.103)	0.144 (0.104)	0.141 (0.103)
7	0.121 (0.100)	0.137 (0.105)	0.133 (0.103)	0.124 (0.103)	0.122 (0.103)

Table 3.3 MML and HB estimates for log-log link

i	MMLE	$\hat{\alpha}_i^{HB} - \hat{\alpha}_8^{HB}$			
		$a = 1$	$a = .5$	$a = 0.01$	$a = 0.001$
1	0.875 (0.132)	0.892 (0.130)	0.850 (0.134)	0.880 (0.131)	0.884 (0.133)
2	2.033 (0.197)	2.065 (0.200)	2.049 (0.195)	2.049 (0.199)	2.050 (0.197)
3	1.568 (0.165)	1.589 (0.164)	1.582 (0.167)	1.580 (0.164)	1.581 (0.163)
4	1.684 (0.171)	1.712 (0.171)	1.697 (0.170)	1.691 (0.171)	1.700 (0.173)
5	-0.471 (0.110)	-0.466 (0.110)	-0.470 (0.108)	-0.470 (0.106)	-0.469 (0.110)
6	0.513 (0.120)	0.528 (0.118)	0.521 (0.121)	0.512 (0.121)	0.520 (0.121)
7	0.535 (0.121)	0.545 (0.117)	0.542 (0.121)	0.538 (0.120)	0.542 (0.120)

CHAPTER 4

A UNIFORM BAYESIAN ANALYSIS OF TWO-PARAMETER ITEM RESPONSE MODELS FOR BINARY DATA

4.1 Introduction

As mentioned in the introduction, two-parameter item response models involve item discrimination parameters in addition to the subject and item difficulty parameters. In its most general form, the discrimination parameters may vary from item to item. Two-parameter item response models are given either by (1.1.2) or by (1.1.3), i.e.,

$$p_{ij} = F(\gamma_j(\theta_i + \alpha_j)), \quad (4.1.1)$$

or alternately as

$$p_{ij} = F(\gamma_j\theta_i + b_j), \quad (4.1.2)$$

This chapter addresses the issue of choice of priors for two-parameter item response models. We have discussed in Chapters 2 and 3 how nonidentifiable posteriors can often be improper. We shall note in this chapter that even if the posterior is identifiable, a choice of improper priors can often lead to improper posteriors.

In Section 4.2, we examine critically the issue of the choice of priors. Much of the existing work employs noninformative priors for one or more of the set of parameters $\{\theta_i\}$, $\{\alpha_j\}$ and $\{\gamma_j\}$. This is especially tempting when there is vague prior information. However, this leads to the possibility of improper posteriors. This impropriety is sometimes due to a nonidentifiable posterior. But, as we show in Section

2, impropriety may result even otherwise. For example, it is shown that the priors employed by Albert (1992) and Kim et al. (1994) lead to identifiable but improper posteriors. In this context, we alert to the possibility that it is not always easy to recognize the impropriety unless one carries out the full integration analytically. For example, in the Markov Chain Monte Carlo (MC²) numerical integration technique used by Albert (1992), all the full conditionals are proper, and yet the joint posterior may be improper. For Kim et al. (1994), the posterior modes are finite, but the joint posterior may still be improper.

Our recommendation, therefore, is to use proper priors throughout for the θ_i, α_j and γ_j . In Section 4.3, we have proposed a general algorithm for implementing the Bayes procedure via MC² with proper priors. In the final Section we have illustrated the application of the proposed methodology with the aid of a real dataset for three important link functions, namely the logit, probit and the log-log.

4.2 Choice Of Priors

Let $\mathbf{x} = (x_{11}, \dots, x_{1k}, \dots, x_{n1}, \dots, x_{nk})$, $\boldsymbol{\theta} = (\theta_1, \dots, \theta_n)$, $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_k)$, $\boldsymbol{\gamma} = (\gamma_1, \dots, \gamma_k)$ and $\mathbf{b} = (b_1, \dots, b_k)$. For a general link function F , based on (4.1.2), the likelihood function is given by

$$L(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\gamma} | \mathbf{x}) = \prod_{i=1}^n \prod_{j=1}^k [F^{x_{ij}}(\gamma_j \theta_i + b_j) \bar{F}^{1-x_{ij}}(\gamma_j \theta_i + b_j)] \quad (4.2.1)$$

A general class of priors for $(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\gamma})$ is given by

$$\pi(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\gamma}) = \left[\prod_{i=1}^n g_{1i}(\theta_i) \right] \left[\prod_{j=1}^k g_{2j}(b_j) \right] \left[\prod_{j=1}^k g_{3j}(\gamma_j) \right] \quad (4.2.2)$$

A basic question here is the selection of priors $\{g_{1i}\}$, $\{g_{2j}\}$ and $\{g_{3j}\}$. Much of the existing literature on this topic recommends using noninformative priors for one

or more of the set of parameters $\{\theta_i\}$, $\{b_j\}$ and $\{\gamma_j\}$. For example, for the two-parameter logistic model, with the formulation given in (4.1.2), Swaminathan and Gifford (1985) recommended using flat priors for the $\{\theta_i\}$ and the $\{b_j\}$ (see their equations (12) and (13), and notice the difference with our notations). Albert (1992) recommended using flat priors for the $\{b_j\}$ and $\{\gamma_j\}$ for the two-parameter probit model with the formulation given in (4.1.2). It turns out that such priors will lead to improper posteriors.

We begin with a theorem which establishes the impropriety of the posterior when at least one of the γ_j is improper.

Theorem 1.

Consider the prior given in (4.1.2), where at least one of the g_{3j} is improper. Then the joint posterior of (θ, b, γ) given $\mathbf{x}$ is improper.

Proof.

The joint posterior of (θ, b, γ) given $\mathbf{x}$ is

$$\begin{aligned}
 \pi(\theta, b, \gamma | \mathbf{x}) &\propto L(\theta, b, \gamma | \mathbf{x}) \left[\prod_{i=1}^n g_{1i}(\theta_i) \right] \left[\prod_{j=1}^k g_{2j}(b_j) \right] \left[\prod_{j=1}^k g_{3j}(\gamma_j) \right] \\
 &= \prod_{i=1}^n \prod_{j=1}^k [F^{x_{ij}}(\gamma_j \theta_i + b_j) \bar{F}^{1-x_{ij}}(\gamma_j \theta_i + b_j)] \left[\prod_{i=1}^n g_{1i}(\theta_i) \right] \times \\
 &\quad \left[\prod_{j=1}^k g_{2j}(b_j) \right] \left[\prod_{j=1}^k g_{3j}(\gamma_j) \right]
 \end{aligned} \tag{4.2.3}$$

Write

$$\begin{aligned}
 I_i &= \int_{-\infty}^{\infty} \prod_{i=1}^n \prod_{j=1}^k [F^{x_{ij}}(\gamma_j \theta_i + b_j) \bar{F}^{1-x_{ij}}(\gamma_j \theta_i + b_j)] \\
 &\quad g_{1i}(\theta_i) d\theta_i, \quad (i = 1, \dots, n).
 \end{aligned} \tag{4.2.4}$$

Then

$$\pi(\boldsymbol{\theta}, \mathbf{b}|\mathbf{x}) \propto \left(\prod_{i=1}^n I_i \right) \prod_{j=1}^k g_{2j}(b_j) \prod_{j=1}^k g_{3j}(\gamma_j). \quad (4.2.5)$$

For definiteness, assume that $g_{31}(\gamma_1)$ is improper. Now for each fixed i , $x_{i1} = 1$ or 0 . Hence, each integral I_i contains a term $F(\gamma_1\theta_i + b_1)$ or $\bar{F}(\gamma_1\theta_i + b_1)$, but not both. In the first case,

$$\begin{aligned} I_i &\geq \int_0^\infty \prod_{j=1}^k [F^{x_{ij}}(\gamma_j\theta_i + b_j)) \bar{F}^{1-x_{ij}}(\gamma_j\theta_i + b_j)] g_{1i}(\theta_i) d\theta_i \\ &\geq F(b_1) \int_0^\infty \prod_{j=2}^k [F^{x_{ij}}(\gamma_j\theta_i + b_j)) \bar{F}^{1-x_{ij}}(\gamma_j\theta_i + b_j)] g_{1i}(\theta_i) d\theta_i \end{aligned} \quad (4.2.6)$$

In the second case,

$$\begin{aligned} I_i &\geq \int_{-\infty}^0 \prod_{j=1}^k [F^{x_{ij}}(\gamma_j\theta_i + b_j)) \bar{F}^{1-x_{ij}}(\gamma_j\theta_i + b_j)] g_{1i}(\theta_i) d\theta_i \\ &\geq \bar{F}(b_1) \int_{-\infty}^0 \prod_{j=2}^k [F^{x_{ij}}(\gamma_j\theta_i + b_j)) \bar{F}^{1-x_{ij}}(\gamma_j\theta_i + b_j)] g_{1i}(\theta_i) d\theta_i \end{aligned} \quad (4.2.7)$$

This shows that we can get a lower bound for $\prod_{i=1}^n I_i$ which is nonnegative, and *does not involve* γ_1 . Integrating this lower bound with respect to the marginal pdf of γ_1 over $(0, \infty)$ one gets

$$\int_0^\infty \left(\prod_{i=1}^n I_i \right) g_{31}(\gamma_1) d\gamma_1 = +\infty$$

This implies that the posterior $\pi(\mathbf{b}, \boldsymbol{\gamma}|\mathbf{x})$ is improper, and hence $\pi(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\gamma}|\mathbf{x})$ is also improper.

Remark 1. Albert (1992) considered a special case where $F = \Phi$, θ_i are iid $N(0, 1)$ and $g_{2j}(b_j) \propto 1$, $g_{3j}(\gamma_j) \propto 1$ for $j = 1, \dots, k$. As a consequence of the above theorem, it follows that Albert's prior leads to an improper posterior.

Remark 2. Note however, that all the full conditionals are proper in this case so that

impropriety of the posterior may not necessarily be detected by the MC² technique. To see this notice that

$$\pi(\theta_i|\theta_l(l \neq i), \mathbf{b}, \boldsymbol{\gamma}, \mathbf{x}) \propto \prod_{j=1}^k [F^{x_{ij}}(\gamma_j \theta_i + b_j)] \bar{F}^{1-x_{ij}}(\gamma_j \theta_i + b_j)] g_{1i}(\theta_i)$$

which is typically integrable with respect to θ_i . Similarly, $\pi(b_j|b_m(m \neq j), \boldsymbol{\theta}, \boldsymbol{\gamma}, \mathbf{x})$, $\pi(\gamma_j|\gamma_m(m \neq j), \boldsymbol{\theta}, \mathbf{b}, \mathbf{x})$ may all be proper. Thus, it is possible to generate samples from these proper full conditionals without detecting the impropriety of the posteriors.

Remark 3. The impropriety of the posterior continues to hold under Albert's (1992) prior with the parametrization given in (4.1.2). To see this all one needs to note is the one-to-oneness of $(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ with $(\boldsymbol{\theta}, \mathbf{b}, \boldsymbol{\gamma})$.

Swaminathan and Gifford (1985, p353) suggested that using flat priors for $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, the Bayesian analysis will be equivalent to a likelihood based analysis. But as we will see now, the Bayesian analysis then becomes questionable because the posterior is improper.

The above fact is a consequence of a more general result which is proved below.

Theorem 2.

Consider the prior $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) = g(\boldsymbol{\gamma})$ for $\boldsymbol{\theta}, \boldsymbol{\alpha}$ and $\boldsymbol{\gamma}$, where g is an arbitrary positive function of $\boldsymbol{\gamma}$. Then the posterior $\pi(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{x})$ is always improper.

Proof.

First write

$$\begin{aligned} \pi(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{x}) &\propto L(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}|\mathbf{x}) \pi(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) \\ &= \prod_{i=1}^n \prod_{j=1}^k [F^{x_{ij}}(\gamma_j(\theta_i + \alpha_j)) \bar{F}^{1-x_{ij}}(\gamma_j(\theta_i + \alpha_j))] g(\boldsymbol{\gamma}) \quad (4.2.8) \end{aligned}$$

With the one-to-one transformation

$$\eta_i = \theta_i + \alpha_k \quad (i = 1, \dots, n),$$

$$\begin{aligned}\xi_j &= \alpha_j - \alpha_k \quad (j = 1, \dots, k-1), \\ \alpha_k &= \alpha_k,\end{aligned}\tag{4.2.9}$$

the joint posterior of $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)$, $\boldsymbol{\xi} = (\xi_1, \dots, \xi_{k-1})$, α_k and $\boldsymbol{\gamma}$ is given by

$$\begin{aligned}\pi(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k, \boldsymbol{\gamma} | \mathbf{x}) &\propto \prod_{i=1}^n \prod_{j=1}^{k-1} [F^{x_{ij}}(\gamma_j(\eta_i + \xi_j)) \bar{F}^{1-x_{ij}}(\gamma_j(\eta_i + \xi_j))] \times \\ &\quad \prod_{i=1}^n [F^{x_{ik}}(\gamma_k \eta_i) \bar{F}^{1-x_{ik}}(\gamma_k \eta_i)] g(\boldsymbol{\gamma}).\end{aligned}\tag{4.2.10}$$

The above posterior does not depend on α_k . Hence, integrating over α_k which ranges from $(-\infty, \infty)$,

$$\int_{-\infty}^{\infty} \pi(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k, \boldsymbol{\gamma} | \mathbf{x}) d\alpha_k = +\infty.\tag{4.2.11}$$

This implies the impropriety of the joint posterior of $(\boldsymbol{\eta}, \boldsymbol{\xi}, \alpha_k, \boldsymbol{\gamma})$ and equivalently of that of $(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$.

Kim et al. (1994) considered a prior for $(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ which is outside the class of priors considered in (4.1.1). However, their prior also leads to an improper posterior. Kim et al. (1994) considered the logit link. As we show now, the impropriety holds under an arbitrary link. We may point out though that Kim et al. (1994) were primarily interested in the issue of joint versus marginal modes. These modes are finite even though the posterior is improper.

Kim et al. consider the following hierarchical prior:

θ_i ($i = 1, \dots, n$), α_j ($j = 1, \dots, k$) and $\beta_j = \log \gamma_j$ ($j = 1, \dots, k$) are mutually independent, and

$$\begin{aligned}\theta_i &\stackrel{iid}{\sim} N(0, 1) \\ \log \gamma_j &= \beta_j | \mu_\beta, \sigma_\beta^2 \stackrel{iid}{\sim} N(\mu_\beta, \sigma_\beta^2)\end{aligned}$$

$$\begin{aligned}
\alpha_j | \mu_\alpha, \sigma_\alpha^2 &\stackrel{iid}{\sim} N(\mu_\alpha, \sigma_\alpha^2) \\
\mu_\alpha \text{ and } \mu_\beta &\text{ are uniform } (-\infty, \infty) \\
r_\alpha &= (\sigma_\alpha^2)^{-1} \sim \text{Gamma}(\nu_\alpha, \lambda_\alpha) \\
r_\beta &= (\sigma_\beta^2)^{-1} \sim \text{Gamma}(\nu_\beta, \lambda_\beta)
\end{aligned} \tag{4.2.12}$$

Theorem 3.

Under the choice of priors given in (4.2.12) the posterior $f(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{x})$ is improper.

Proof.

Note that the marginal priors for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ after integrating out with respect to the nuisance parameters $\mu_\alpha, \mu_\beta, r_\alpha$ and r_β are given by,

$$\begin{aligned}
\pi(\boldsymbol{\alpha} | \nu_\alpha, \lambda_\alpha) &\propto \left\{ \frac{2}{\lambda_\alpha} + \sum_{j=1}^k (\alpha_j - \bar{\alpha})^2 \right\}^{-(k+2\nu_\alpha-1)/2} \\
\pi(\boldsymbol{\beta} | \nu_\beta, \lambda_\beta) &\propto \left\{ \frac{2}{\lambda_\beta} + \sum_{j=1}^k (\beta_j - \bar{\beta})^2 \right\}^{-(k+2\nu_\beta-1)/2}
\end{aligned} \tag{4.2.13}$$

Then the joint posterior of $\boldsymbol{\theta}, \boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is given by,

$$\begin{aligned}
f(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\beta} | \mathbf{x}) &\propto \prod_{i=1}^n \prod_{j=1}^k \left[F^{x_{ij}}(\exp(\beta_j)(\theta_i + \alpha_j)) \bar{F}^{1-x_{ij}}(\exp(\beta_j)(\theta_i + \alpha_j)) \right] \times \\
&\exp\left(-\frac{1}{2} \sum_{i=1}^n \theta_i^2\right) \left\{ \frac{2}{\lambda_\alpha} + \sum_{j=1}^k (\alpha_j - \bar{\alpha})^2 \right\}^{-(k+2\nu_\alpha-1)/2} \times \\
&\left\{ \frac{2}{\lambda_\beta} + \sum_{j=1}^k (\beta_j - \bar{\beta})^2 \right\}^{-(k+2\nu_\beta-1)/2}
\end{aligned} \tag{4.2.14}$$

Use the transformation $\zeta_j = \beta_j - \beta_k$ ($j = 1, \dots, k-1$) and let $\bar{\zeta} = k^{-1} \sum_{j=1}^{k-1} \zeta_j$. Then the joint posterior of $\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_{k-1})$ and β_k is given by

$$f(\theta, \alpha, \zeta, \beta_k | \mathbf{x}) \propto$$

$$\begin{aligned} & \prod_{i=1}^n \prod_{j=1}^{k-1} \left[F^{x_{ij}}(\exp(\zeta_j + \beta_k)(\theta_i + \alpha_j)) \bar{F}^{1-x_{ij}}(\exp(\zeta_j + \beta_k)(\theta_i + \alpha_j)) \right] \times \\ & \prod_{i=1}^n \left[F^{x_{ik}}(\exp(\beta_k)(\theta_i + \alpha_k)) \bar{F}^{1-x_{ik}}(\exp(\beta_k)(\theta_i + \alpha_k)) \right] \times \\ & \exp\left(-\frac{1}{2} \sum_{i=1}^n \theta_i^2\right) \left\{ \frac{2}{\lambda_\alpha} + \sum_{j=1}^k (\alpha_j - \bar{\alpha})^2 \right\}^{-(k+2\nu_\alpha-1)/2} \times \\ & \left\{ \frac{2}{\lambda_\beta} + \sum_{j=1}^{k-1} (\zeta_j - \bar{\zeta})^2 \right\}^{-(k+2\nu_\beta-1)/2} \end{aligned} \quad (4.2.15)$$

Note that for $\theta_i \geq 0$ ($i = 1, \dots, n$), $\alpha_j \geq 0$ ($j = 1, \dots, k$)

$$\frac{F(\exp(\zeta_j + \beta_k)(\theta_i + \alpha_j))}{\bar{F}(\exp(\zeta_j + \beta_k)(\theta_i + \alpha_j))} \geq \frac{F(0)}{\bar{F}(0)}.$$

Hence, for this set of parameter values and $\beta_k \leq 0$, the right hand side of (4.2.15)

$$\begin{aligned} & \geq \prod_{i=1}^n \prod_{j=1}^{k-1} \bar{F}(\exp(\zeta_j + \beta_k)(\theta_i + \alpha_j)) \prod_{i=1}^n \bar{F}(\exp(\beta_k)(\theta_i + \alpha_k)) \times \\ & \exp\left(-\frac{1}{2} \sum_{i=1}^n \theta_i^2\right) \left\{ \frac{2}{\lambda_\alpha} + \sum_{j=1}^k (\alpha_j - \bar{\alpha})^2 \right\}^{-(k+2\nu_\alpha-1)/2} \times \\ & \left\{ \frac{2}{\lambda_\beta} + \sum_{j=1}^{k-1} (\zeta_j - \bar{\zeta})^2 \right\}^{-(k+2\nu_\beta-1)/2} \\ & \geq \prod_{i=1}^n \prod_{j=1}^{k-1} \bar{F}(\exp(\zeta_j)(\theta_i + \alpha_j)) \prod_{i=1}^n \bar{F}(\theta_i + \alpha_k) \times \\ & \exp\left(-\frac{1}{2} \sum_{i=1}^n \theta_i^2\right) \left\{ \frac{2}{\lambda_\alpha} + \sum_{j=1}^k (\alpha_j - \bar{\alpha})^2 \right\}^{-(k+2\nu_\alpha-1)/2} \times \end{aligned}$$

$$\left\{ \frac{2}{\lambda_\beta} + \sum_{j=1}^{k-1} (\zeta_j - \bar{\zeta})^2 \right\}^{-(k+2\nu_\beta-1)/2} \quad (4.2.16)$$

The right hand side of (4.2.16) does not depend on β_k . Integrating this with respect to β_k over $(-\infty, 0)$, it follows that

$$\int \cdots \int f(\boldsymbol{\theta}, \boldsymbol{\alpha}, \boldsymbol{\zeta}, \beta_k | \mathbf{x}) d\boldsymbol{\theta} d\boldsymbol{\alpha} d\boldsymbol{\zeta} d\beta_k = +\infty.$$

This proves the theorem.

The above theorems show that noninformative priors for problems of this type often lead to improper posteriors. We have therefore, used proper priors in the remainder of this chapter, and in the analysis of the dataset given in the earlier chapter.

4.3 Implementation Of Bayes Procedures

Consider the general link $p_{ij} = F(\gamma_j \eta_i + b_j)$, $i = 1, \dots, n$, $j = 1, \dots, k$, and the prior $g(\boldsymbol{\eta}, \mathbf{b}, \boldsymbol{\gamma}) \propto \prod_{i=1}^n g_1(\eta_i) \prod_{j=1}^k g_2(b_j) g_3(\gamma_j)$. Due to nonconjugacy of the prior, the posterior is analytically intractable, and can be found only via numerical integration. Also, direct numerical integration seems infeasible in this case because of the high dimensionality of the problem.

Fortunately, the integration task has become easier due to the advent of the sophisticated Markov Chain Monte Carlo (MC²) numerical integration techniques. We shall use, as before, Gibbs sampling for integration purposes. In this case the full conditionals are given by

$$f(\eta_i | \eta_l (l \neq i), \mathbf{b}, \boldsymbol{\gamma}, \mathbf{x}) \propto \prod_{j=1}^{k-1} [F^{x_{ij}}(\gamma_j \eta_i + b_j) \bar{F}^{1-x_{ij}}(\gamma_j \eta_i + b_j)] g_1(\eta_i); \quad (4.3.1)$$

$$f(b_j | b_m (m \neq j), \boldsymbol{\eta}, \boldsymbol{\gamma}, \mathbf{x}) \propto \prod_{i=1}^n [F^{x_{ij}}(\gamma_j \eta_i + b_j) \bar{F}^{1-x_{ij}}(\gamma_j \eta_i + b_j)] g_2(b_j); \quad (4.3.2)$$

and

$$f(\gamma_j | \gamma_s (s \neq j), \boldsymbol{\eta}, \mathbf{b}, \mathbf{x}) \propto \prod_{i=1}^n [F^{x_{ij}}(\gamma_j \eta_i + b_j) \bar{F}^{1-x_{ij}}(\gamma_j \eta_i + b_j)] g_3(\gamma_j); \quad (4.3.3)$$

$i = 1, \dots, n$; $j = 1, \dots, k$. Note that the full conditional of η_i does not involve the remaining $\eta_l (l \neq i)$. Similar, the full conditional of b_j does not involve the remaining $b_m (m \neq j)$. A similar remark holds for the full conditional of γ_j . Notice, however that the full conditionals are typically non-standard densities from which it is not possible to draw samples directly. The general procedure for generating samples in such cases is to use the Metropolis-Hastings accept-reject algorithm. If, however, F , $\bar{F}$, g_1 , g_2 and g_3 are log-concave, the full conditionals $f(\eta_i | \cdot)$, $f(b_j | \cdot)$ and $f(\gamma_j | \cdot)$ are all log-concave. We can then use the Adaptive Rejection sampling of Gilks and Wild (1992).

Theorem 4.

If F is an IFR distribution function and g_1 , g_2 and g_3 are log-concave densities then the full conditionals, $f(\eta_i | \cdot)$, $f(b_j | \cdot)$ and $f(\gamma_j | \cdot)$ are all log-concave.

Proof.

To see the log-concavity of $f(\eta_i | \cdot)$, $f(b_j | \cdot)$ and $f(\gamma_j | \cdot)$, we write

$$\log f(\eta_i | \eta_l (l \neq i), \mathbf{b}, \boldsymbol{\gamma}, \mathbf{x}) = \sum_{j=1}^{k-1} [x_{ij} \log F(\gamma_j \eta_i + b_j) + (1 - x_{ij}) \log \bar{F}(\gamma_j \eta_i + b_j)] + \log g_1(\eta_i) \quad (4.3.4)$$

If F , $\bar{F}$ and g_1 are all log-concave then clearly $f(\eta_i | \eta_l (l \neq i), \mathbf{b}, \boldsymbol{\gamma}, \mathbf{x})$ is log-concave.

Similarly,

$$\begin{aligned} \log f(b_j | \boldsymbol{\eta}, b_m (m \neq j), \boldsymbol{\gamma}, \mathbf{x}) &= \sum_{i=1}^n [x_{ij} \log F(\gamma_j \eta_i + b_j) + (1 - x_{ij}) \log \bar{F}(\gamma_j \eta_i + b_j)] \\ &+ \log g_2(b_j) \end{aligned} \quad (4.3.5)$$

and

$$\log f(\gamma_j|\boldsymbol{\eta}, \mathbf{b}, \gamma_s(s \neq j), \mathbf{x}) = \sum_{i=1}^n [x_{ij} \log F(\gamma_j \eta_i + b_j) + (1 - x_{ij}) \log \bar{F}(\gamma_j \eta_i + b_j)] + \log g_3(\gamma_j) \quad (4.3.6)$$

Hence, if F , $\bar{F}$, g_2 and g_3 are log-concave, $f(b_j|\boldsymbol{\eta}, b_m(m \neq j), \boldsymbol{\gamma}, \mathbf{x})$ and $f(\gamma_j|\gamma_s(s \neq j), \boldsymbol{\eta}, \mathbf{b}, \mathbf{x})$ are log-concave. The log-concavity of F and $\bar{F}$ is ensured since F is a IFR df.

4.4 An Example

In this section we consider the same placement data that we have been considering in the previous two chapters. Our interest lies in inference about estimating the item parameters and the discrimination parameters, or more specifically in the posterior means and posterior s.d.'s of b_j and γ_j ($j = 1, \dots, 7$). Three different links, logit, probit and log-log are considered.

Suppose $\boldsymbol{\eta}$ are iid normal, as is chosen in Swaminathan and Gifford (1985). Specifically we take $N(0, 1)$ distributions for $\boldsymbol{\eta}$. The normality assumption for ability parameters is convenient and may seem reasonable because previous studies may suggest that the population of abilities of subjects is bell shaped and the subjects taking the test are a random sample from this population. If, on the other hand, a test uses a battery of questions for the first time, there may be little (if any) information about the distribution of the question difficulty and discrimination parameters. In that case, a flat or highly diffuse prior may be sensible for them. Now we have seen that if we take a flat prior for both item parameters, then the posterior is improper. Now if an item bank is available, then the normality assumption for the item parameters appears to be reasonable (Lord and Novick, 1968), because one can assume that the items included in the test are coming from a bell shaped population of items. So, we

take $\mathbf{b}$ as iid normal. Swaminathan and Gifford (1985) used a chi-distribution on the discriminating parameters. We put a half normal distribution on the discrimination parameters γ . The entire prior specification is listed below.

$$\begin{aligned}\eta_i &\stackrel{iid}{\sim} N(0, 1) \\ b_j &\stackrel{iid}{\sim} N(0, \tau^2) \\ \gamma_j &\stackrel{iid}{\sim} N(0, \sigma^2) I_{[\gamma > 0]}.\end{aligned}\tag{4.4.1}$$

We take $\tau^2 = 1000$ to allow as vague a prior for $\mathbf{b}$ as possible. For the Bayesian analysis, σ is taken as the tuning parameter and is used to study the sensitivity of the Bayes procedure. The posterior means and s.d.'s are calculated using the Gibbs sampler (Geman and Geman, 1984, Gelfand and Smith, 1990). For this example, the number of iterates to generate a sample is taken as 50, while the number of samples is taken as 5000.

The usual competitors of the Bayes estimates of γ and $\mathbf{b}$ are the maximum likelihood estimates for the mixed model treating θ as a random effect. The associated standard errors for the item parameters are based on the asymptotic distribution of the marginal maximum likelihood estimators out of bounds. (MMLE). A criticism against maximum likelihood approach is that, a limit should be imposed on the range of values taken by the discrimination parameters. This is actually tantamount to the specification of prior belief without a Bayesian justification.

A natural question in regard to applying Bayesian technique is the selection of σ , the standard deviation of the distribution of the discrimination parameters. If a value of σ is available from previous work, that value can be used in these calculations. However, in order to select a value for σ , a 95% credibility interval for γ_j may be helpful. The mean for the distribution of γ_j is 0 and the variance is $\sigma^2/4$. Then an approximate 95% credibility interval for γ_j is given by, $\pm z_{0.025} \sigma/2$. Hence, the width of the interval is, say $W = \sigma z_{0.025}$, where $z_{0.025}$ represents the upper 0.025 percentage

point of the $N(0, 1)$ distribution. Thus, $\sigma = (W/z_{0.025})$. The following table (Table 4.1) provides some values of σ for different values of W . Thus, with a specification of the width W for an interval of γ_j , it is possible to specify the prior distribution of γ_j . Now, the information provided by the credibility intervals for which W is in $(1, 5)$ is not too precise and not too vague. Hence, one way to choose σ is the solution of the above equation with W in $(1, 5)$. We used three values of σ , namely, 0.5, 1.0 and 1.5.

Table 4.1 Width of the 95% credibility interval for γ_j and the corresponding σ

W	σ
1	0.51
1.5	0.77
2	1.02
2.5	1.28
3	1.53
5	2.55
10	5.1

We report the marginal maximum likelihood estimates of $\mathbf{b}$ and $\boldsymbol{\gamma}$ and compare them with the corresponding Bayes estimates. The estimates for the three link functions are provided in Tables (4.2)-(4.4). Table 4.2 displays the Bayes estimates of the item effects, with the associated standard errors in parentheses, for the logit link using $\sigma = .50, 1.0$, and 1.5 . Tables 4.3 and 4.4 show the Bayes estimates for the probit and log-log links. All the three tables also provide the marginal MLEs. The marginal MLEs are also the posterior modes for the Bayesian approach which uses flat prior for $(\mathbf{b}, \boldsymbol{\gamma})$. So not surprisingly, some of the marginal MLEs for $\mathbf{b}$ are very close to the Bayes estimates and the MMLEs of $\boldsymbol{\gamma}$ are also somewhat close to the Bayes estimates. We also notice that the Bayes estimates tend to increase as the tuning parameter σ increases. In the previous chapters we observed the same type of phenomenon. An obvious question is how to judge goodness of fit for this model. We need some measure to check whether this model is giving a better fit than the simpler

model, i.e., the one-parameter model. This issue of model goodness of fit, which can be a topic of future research, was not considered in this dissertation.

Table 4.2 MML and Bayes estimates for logit link

i	MMLE		Bayes Estimates					
	b	γ	$\sigma = 0.5$		$\sigma = 1.0$		$\sigma = 1.5$	
			b	γ	b	γ	b	γ
1	0.906 (0.239)	1.222 (0.362)	0.877 (0.221)	1.136 (0.303)	0.915 (0.239)	1.234 (0.354)	0.927 (0.247)	1.256 (0.359)
2	2.301 (0.383)	1.104 (0.392)	2.223 (0.323)	0.958 (0.313)	2.300 (0.357)	1.062 (0.356)	2.341 (0.388)	1.118 (0.392)
3	2.079 (0.415)	1.544 (0.469)	1.903 (0.326)	1.269 (0.343)	1.995 (0.354)	1.409 (0.386)	2.071 (0.393)	1.514 (0.432)
4	1.470 (0.209)	0.428 (0.242)	1.481 (0.204)	0.421 (0.209)	1.497 (0.215)	0.447 (0.229)	1.505 (0.212)	0.464 (0.227)
5	-1.272 (0.211)	1.077 (0.339)	-1.268 (0.206)	0.946 (0.287)	-1.277 (0.211)	1.027 (0.319)	-1.287 (0.217)	1.072 (0.340)
6	0.469 (0.231)	1.555 (0.436)	0.416 (0.213)	1.345 (0.327)	0.443 (0.223)	1.460 (0.367)	0.461 (0.231)	1.530 (0.417)
7	0.181 (0.158)	0.531 (0.212)	0.181 (0.159)	0.529 (0.207)	0.183 (0.157)	0.537 (0.208)	0.186 (0.160)	0.545 (0.212)

Table 4.3 MML and Bayes estimates for probit link

i	MMLE		Bayes Estimates					
	b	γ	$\sigma = 0.276$		$\sigma = 0.551$		$\sigma = 0.827$	
			b	γ	b	γ	b	γ
1	0.604 (0.143)	0.747 (0.202)	0.594 (0.147)	0.852 (0.227)	0.605 (0.151)	0.899 (0.245)	0.602 (0.151)	0.888 (0.251)
2	1.321 (0.186)	0.563 (0.201)	1.304 (0.181)	0.632 (0.224)	1.327 (0.162)	0.673 (0.209)	1.338 (0.206)	0.691 (0.263)
3	1.270 (0.229)	0.899 (0.259)	1.240 (0.215)	1.004 (0.272)	1.255 (0.177)	1.040 (0.229)	1.306 (0.241)	1.109 (0.305)
4	0.912 (0.122)	0.265 (0.143)	0.919 (0.119)	0.327 (0.159)	0.897 (0.120)	0.294 (0.148)	0.922 (0.123)	0.334 (0.166)
5	-0.692 (0.112)	0.606 (0.179)	-0.713 (0.114)	0.696 (0.195)	-0.710 (0.116)	0.780 (0.192)	-0.713 (0.113)	0.719 (0.208)
6	0.370 (0.145)	0.958 (0.245)	0.346 (0.137)	1.067 (0.239)	0.358 (0.147)	1.138 (0.253)	0.360 (0.146)	1.127 (0.273)
7	0.146 (0.101)	0.353 (0.131)	0.137 (0.100)	0.401 (0.150)	0.136 (0.100)	0.424 (0.153)	0.142 (0.103)	0.423 (0.157)

Table 4.4 MML and Bayes estimates for log-log link

i	MMLE		Bayes Estimates					
	b	γ	$\sigma = 0.354$		$\sigma = 0.707$		$\sigma = 1.061$	
			b	γ	b	γ	b	γ
1	1.077 (0.195)	0.946 (0.272)	1.054 (0.181)	0.899 (0.244)	1.097 (0.195)	0.978 (0.279)	1.094 (0.197)	0.983 (0.283)
2	2.303 (0.342)	1.00 (0.341)	2.235 (0.293)	0.879 (0.292)	2.307 (0.317)	0.975 (0.322)	2.310 (0.335)	0.984 (0.346)
3	2.01 (0.324)	1.218 (0.347)	1.917 (0.271)	1.092 (0.280)	1.990 (0.322)	1.180 (0.343)	2.005 (0.312)	1.214 (0.341)
4	1.572 (0.187)	0.407 (0.211)	1.589 (0.183)	0.392 (0.191)	1.597 (0.185)	0.409 (0.201)	1.592 (0.189)	0.404 (0.201)
5	-0.380 (0.101)	0.526 (0.167)	-0.384 (0.105)	0.600 (0.174)	-0.381 (0.104)	0.559 (0.180)	-0.388 (0.105)	0.580 (0.195)
6	0.702 (0.165)	0.971 (0.278)	0.692 (0.157)	0.961 (0.247)	0.724 (0.171)	1.041 (0.316)	0.735 (0.182)	1.082 (0.355)
7	0.501 (0.119)	0.399 (0.149)	0.508 (0.119)	0.400 (0.151)	0.511 (0.121)	0.409 (0.150)	0.508 (0.117)	0.404 (0.153)

CHAPTER 5

SUMMARY AND FUTURE RESEARCH

This dissertation primarily focuses on developing Bayesian methods for one- and two-parameter item response models. Most of the Bayesian articles in the literature have so far been link specific. We brought many of the results together under a general link function. One of the main objectives was to develop Bayesian methods which ensured propriety of the posteriors. We discussed extensively throughout this dissertation the choice of priors that leads to proper posterior distributions. We have provided a general algorithm to implement our methodology in a given context.

In Chapter 2, we developed a Bayesian technique for one-parameter item response models for general link function. We derived specific conditions for the posteriors to be proper while using improper priors. We also talked about marginal maximum likelihood estimates for item parameters and provided conditions under which they are consistent estimators. Also a uniform approximation of improper priors by proper ones is provided in this chapter. We separately deal with Bayesian estimation of item parameters in matched pairs case. One important issue was to see the influence (if any) of the main diagonal counts on the Bayes estimates. We have shown that unlike other methods, such as conditional and marginal maximum likelihood, the Bayes procedure does depend on the main diagonal counts for 2×2 tables.

In Chapter 3 we developed a hierarchical Bayes method for one-parameter item response models. Here we have provided conditions which lead to proper posteriors for item parameters. We discussed a general algorithm for implementation of the procedure in any given situation.

In Chapter 4 we give a classical Bayesian methodology for two-parameter item response models. We notice very carefully the problem of improper posteriors and suggested using proper priors only. A general algorithm is also provided to implement the technique.

Bayes procedure has a few advantages over the marginal maximum likelihood approach. An obvious advantage is that, with a hierarchical prior one can avoid choosing an arbitrary value for the scale parameter of the subject ability distribution. Another criticism against maximum likelihood approach is that, for two-parameter models a limit should be imposed on the range of values taken by the discrimination parameters. This is actually tantamount to the specification of prior belief without a Bayesian justification.

In all the chapters we use data from a mathematics placement test, obtained from Professor James Albert of Bowling Green State University. Also, throughout the dissertation, the implementation of the Bayes methodology has been illustrated by adopting a Markov Chain Monte Carlo integration technique known as the Gibbs sampler. Using this procedure, the posterior density as well as conditional mean and variance can be obtained with considerable ease. Also, a special technique called adaptive rejection sampling has been extensively used to generate samples from log-concave densities.

As for future research, clearly there are many important questions which are left unanswered. One important consideration is the choice of link functions. One way to answer this is to find the Bayes factors of one model with respect to another, and use the same for model selection. Another important consideration is whether or not to use an adaptive link function, by estimating the link from the data as is done in Mallick and Gelfand (1994), and use a nonparametric procedure eventually.

As mentioned in Chapter 4, an interesting topic of future research could be model selection for a given link. For example, it might be important to know when it is

necessary to use a more complex model with discrimination parameters. Also, a general model goodness of fit measure might help deciding which model to consider in a given situation. This will be a future research topic.

In Chapter 2, Section 3 we have proved a result showing propriety of the posteriors using flat priors for both θ and α for the special case when $k = 2$. In future one should investigate whether this result holds for arbitrary k . In this chapter we also noticed a conflicting behavior in Bayes estimates for matched-pairs case. One should also investigate this in further detail in future.

The general techniques used in this dissertation have the potential of being adapted under more complex parametrizations. A further complexity in these models can be introduced by including a guessing parameter. Such models, usually referred to as three-parameter item response models, will also be a topic of future study.

BIBLIOGRAPHY

- Albert, J.H. (1992). Bayesian estimation of normal ogive item response curve using Gibbs sampling. *Journal of Educational Statistics*, 17, 251-269.
- Andersen, E.B. (1970). Asymptotic properties of conditional maximum likelihood estimators. *The Journal of the Royal Statistical Society, Series B*, 32, 283-301.
- Andersen, E.B. (1972). The numerical solution of a set of conditional estimation equations. *The Journal of the Royal Statistical Society, Series B*, 34, 42-54.
- Andersen, E.B. (1973). Conditional inference for multiple-choice questionnaires. *British Journal of Mathematical and Statistical Psychology*, 26, 31-44.
- Birnbaum, A. (1969). Statistical theory for logistic mental test models with a prior distribution of ability. *Journal of Mathematical Psychology*, 6, 258-276.
- Bock, R.D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: application of an EM algorithm. *Psychometrika*, 46, 443-459.
- Bock, R.D. & Lieberman, M. (1970). Fitting a response model for n dichotomously scored items. *Psychometrika*, 35, 179-197.
- Casella, G. & George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46, 167-174.
- Cox, D.R. (1958). The regression analysis of binary sequences. *J. R. Statist. Soc., B*, 20, 215-242.
- Cox, D.R. & Snell, E.J. (1989). The Analysis of Binary Data. 2nd edn., London: Chapman and Hall.
- Dawid, A.P. (1979). Conditional independence in statistical theory (with discussion). *J. R. Statist. Soc., B*, 41, 1-31.
- De Leeuw, J. & Verhelst, N. (1986). Maximum likelihood estimation in generalized Rasch models. *Journal of Educational and Behavioral Statistics*, 11, 183-196.

- Fischer, G.H. & Molenaar, I.W. (1995). Rasch Models: Foundations, Recent Developments and Applications. New York: Springer-Verlag.
- Follman, D.A. (1988). Consistent estimation in the Rasch model based on nonparametric margins. *Psychometrika*, 53, 553-562.
- Foutz, R.V. (1977). On the unique consistent solution to the likelihood equations. *Journal of the American Statistical Association*, 72, 147-148.
- Gelman, A. & Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-511.
- Ghosh, M. (1995). Inconsistent maximum likelihood estimators for the Rasch model. *Statistics and Probability Letters*, 23, 165-170.
- Gilks, W.R. & Wild, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics*, 41, 337-348.
- Hambleton, R.K. & Swaminathan, H. (1985). Item Response Theory: Principles and Applications. Boston: Kluwer-Nijhoff.
- Kim, S., Cohen, A.S., Baker, F.B., Subkoviak, M.J., and Leonard, T. (1994). An investigation of hierarchical Bayes procedures in item response theory. *Psychometrika*, 59, 405-421.
- Leonard, T., and Novick, M.R. (1985). Bayesian inference and diagnostics for the three parameter logistic model (ONR Technical Report 85-5). Iowa City, IA: The University of Iowa, CADA Research Group.
- Liang, K. & Zeger, S.L. (1988). On the use of Concordant pairs in Matched case-control studies. *Biometrics*, 44, 1145-1156.
- Lindsay, B., Clogg, C.C., & Grego, J. (1991). Semiparametric estimation in the Rasch model and related exponential response models, including a simple latent class model for item analysis. *Journal of the American Statistical Association*, 86, 96-107.
- Lord, F.M. (1952). A theory of test scores. *Psychometric monograph*, 7.
- Lord, F.M. (1953a). An application of confidence intervals and of maximum likelihood to the estimation of an examinee's ability. *Psychometrika*, 18, 57-75.
- Lord, F.M. (1953b). The relation of test score to the trait underlying the test. *Educational and Psychological measurement*, 13, 517-548.
- Mallick, B.K. & Gelfand, A.E. (1994). Generalized linear models with unknown link functions. *Biometrika*, 81, 237-245.

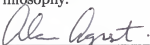
- Mislevy, R.J. & Bock, R.D. (1984). BILOG maximum likelihood item analysis and test scoring: Logistic model. Mooresville, ID: Scientific Software.
- Neuhaus, J.M., Kalbfleisch, J.D., and Hauck, W.W. (1991). A comparison of cluster-specific and population-averaged approaches for analyzing correlated binary data. *International Statistical Review*, 59, 25-35.
- Neyman, J., & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrika*, 16, 1-32.
- O'Hagan, A. (1994). Kendall's Advanced Theory of Statistics. Volume 2B. New York: John Wiley.
- Owen, R. (1975). A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. *Journal of the American Statistical Association*, 70, 351-356.
- Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.
- Rasch, G. (1961). On general laws and the meaning of measurement in psychology. *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics*. Berkeley: University of California press, 4.
- Richardson, M.W. (1936). The relationship between difficulty and differential validity of a test. *Psychometrika*, 1, 33-49.
- Rigdon, S.E. & Tsutakawa, R.K. (1983). Parameter estimation in latent trait models. *Psychometrika*, 48, 567-574.
- Rigdon, S.E. & Tsutakawa, R.K. (1987). Estimation for the Rasch model when both ability and difficulty parameters are random. *Journal of Educational Statistics*, 12, 76-86.
- Sahu, S.K. & Gelfand, A.E. (1995). On propriety of posteriors and Bayesian identifiability in generalized linear models. Preprint.
- Swaminathan, H. & Gifford, J.A. (1981). Estimation of parameters in the three-parameter latent trait model. *Laboratory of Psychometric and Evaluative Research Report No. 93*. Amherst, Mass.: School of Education, University of Massachusetts.
- Swaminathan, H. & Gifford, J.A. (1982). Bayesian estimation in the Rasch model. *Journal of Educational Statistics*, 7, 175-192.
- Swaminathan, H. & Gifford, J.A. (1985). Bayesian estimation in the two-parameter logistic model. *Psychometrika*, 50, 349-364.

- Tsutakawa, R.K. (1984). Estimation of two-parameter logistic item response curves. *Journal of Educational Statistics*, 9, 263-276.
- Tsutakawa, R.K. & Lin, H.Y. (1986). Bayesian estimation of item response curves. *Psychometrika*, 51, 251-267.
- Tsutakawa, R.K. & Soltys, M.J. (1988). Approximation for Bayesian ability estimation. *Journal of Educational Statistics*, 13, 117-130.
- Tsutakawa, R.K. & Johnson, J.C. (1990). The effect of uncertainty of item parameter estimation on ability estimates. *Psychometrika*, 55, 371-390.
- Tucker, L.R. (1946). Maximum validity of a test with equivalent items. *Psychometrika*, 11, 1-13.
-

BIOGRAPHICAL SKETCH

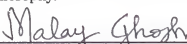
Atalanta Ghosh was born on July 14, 1964, in Calcutta, West Bengal, India. He received his Bachelor of Science degree with statistics major in 1987 from Asutosh College, Calcutta, India. In 1987, he joined University of Calcutta to pursue his Master of Science in statistics. He received this degree in 1989 from University of Calcutta, Calcutta, India. In August 1990 he joined the doctoral program in Statistics at the University of Florida, Gainesville. He expects to receive a Ph.D. degree in December 1996. During his time at the University of Florida, he worked as a research assistant to Dr. Alan Agresti and Dr. Malay Ghosh. He was also employed as a teaching assistant at the Department of Statistics and taught some of the undergraduate courses. He worked as a consultant statistician at the Institute of Food and Agrecultural Sciences, a division of the Department of Statistics under the supervision of Dr. Kenneth M. Portier. Upon graduation, he will join the Computer Task Group Inc. as a consultant Statistician and work as a contractor to the Division of Mathematical Sciences, Eli Lilly, Indianapolis, Indiana.

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



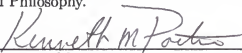
Alan Agresti, Chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Malay Ghosh, Co-chairman
Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



Kenneth M. Portier
Associate Professor of Statistics

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.



James J. Algina
Professor of Foundations of Education

This dissertation was submitted to the Graduate Faculty of the Department of Statistics in the College of Liberal Arts and Sciences and to the Graduate School and was accepted as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

December 1996

Dean, Graduate School